

## Cenários Sintéticos de Vazões para Grandes Sistemas Hídricos Através de Modelos Contemporâneos e Amostragem

Daniel Henrique Marco Detzel<sup>1</sup>, Miriam Rita Moro Mine<sup>2</sup>, Marcelo Bessa<sup>1</sup>, Márcio Bloot<sup>3</sup>

daniel@lactec.org.br; mrmine.dhs@ufpr.br; bessa@lactec.org.br; mlbloot@copel.com

Recebido: 14/01/12 - revisado: 12/1/13 - aceito: 12/9/13

---

### RESUMO

A modelagem de grandes sistemas hídricos mostra-se desafiadora à medida que se aumentam a complexidade e o número de variáveis envolvidas. O caminho comumente encontrado é empregar simplificações como, por exemplo, agregação de reservatórios que operam em uma mesma região. No entanto, estas simplificações podem omitir características importantes do sistema, tornando-o pouco flexível e colocando em risco a confiabilidade dos resultados. No presente artigo, apresenta-se um método alternativo para modelagem de séries de vazões, baseado em formulação autorregressiva com médias móveis contemporânea (CARMA). O modelo é aplicado com sucesso às vazões mensais das 146 usinas hidrelétricas que compõem o Sistema Interligado Nacional, sendo capaz de reproduzir estatísticas de curto e longo termos, além de preservar as correlações espaciais entre as usinas. Ademais, propõe-se a redução do número de séries sintéticas geradas pelo modelo estocástico aplicando-se uma técnica híbrida de classificação de séries e amostragem estratificada não equiprovável. A intenção é viabilizar o uso das séries sintéticas geradas em estudos que exijam maior processamento computacional. Mostra-se que é possível reduzir o número de cenários sintéticos sem distorcer a distribuição de probabilidades obtidas com o modelo estocástico.

**Palavras-Chave:** modelo contemporâneo, séries sintéticas mensais, grandes sistemas hídricos, amostragem.

---

### INTRODUÇÃO

Modelos estocásticos para geração de séries sintéticas de vazões tiveram uso crescente a partir do momento em que se percebeu que a série histórica sozinha era insuficiente para o planejamento apropriado de sistemas de recursos hídricos (JACKSON, 1975). Desde os estudos pioneiros de Thomas & Fiering (1962) e Matalas (1967), diversas técnicas de geração foram desenvolvidas e aplicadas a séries de vazões de diferentes escalas temporais. Em comum, a grande maioria se baseia em métodos clássicos do tipo Box & Jenkins (ou modelos ARIMA, BOX *et al.*, 1994), a fim de modelar apropriadamente a estrutura de correlações inerente às afluições de um rio. Encontram-se também modelos baseados em redes neurais artificiais (CAN *et al.*, 2012), inferência fuzzy (LUNA *et al.*, 2011), wavelets (WANG *et al.*, 2010; NIU *et al.*, 2013) e outros métodos não paramétricos (SALAS & LEE, 2010; HAO & SINGH, 2011).

Em se tratando de séries mensais, a sazonalidade é um elemento extra de grande importância a ser considerado. Para tanto, a geração pode ser feita utilizando modelos do tipo PARMA ou SARI-MA, que contam com um grupo de parâmetros específicos para representação da sazonalidade (MONDAL & WASIMI, 2006; HALTINER & SALAS, 1988; RASMUSSEN *et al.*, 1996; ŞEN, 1978; LUNGU & SEFE, 1991). Alternativamente, podem-se gerar séries anuais e submetê-las a métodos de desagregação (VALENCIA & SCHAAKE, 1973; MEJIA & ROUSSELLE, 1976; STEDINGER *et al.*, 1985; PEREIRA *et al.*, 1984; MAHEEPALA & PERERA, 1996; GUIMARÃES & SANTOS, 2011; SILVA & PORTELA, 2013). Independentemente da técnica escolhida, o uso dos modelos Box & Jenkins possui eficiência comprovada na modelagem de vazões mensais (CASTELLANO-MÉNDEZ *et al.*, 2004).

O procedimento iterativo identificação-estimação-validação proposto por Box & Jenkins (BOX *et al.*, 1994) para ajuste de modelos ARIMA é amplamente difundido na literatura especializada. Entretanto, particularidades de cada estudo, principalmente em relação às suas dimensões, podem fazer com que esse procedimento se torne moroso (embora não menos eficiente). Cita-se como exemplo o Sistema Interligado Nacional (SIN), no qual

---

<sup>1</sup> Instituto de Tecnologia para o Desenvolvimento/LACTEC

<sup>2</sup> Universidade Federal do Paraná/UFPR

<sup>3</sup> Companhia Paranaense de Energia/COPEL

em torno de 85% da energia provém de usinas hidrelétricas. Pode-se dizer que a geração de energia brasileira está sujeita às variações sazonais de diferentes regiões hidrológicas, agregando elevado grau de incerteza ao planejamento do sistema e, conseqüentemente, justificando seu estudo.

Na intenção de contornar o problema da dimensão e complexidade do SIN, o planejamento de médio-longo prazo (horizonte de 60 meses) é feito através de uma técnica de agregação de reservatórios (ARVANITIDIS & ROSING, 1970; CRUZ JR. & SOARES, 1996), na qual usinas instaladas em uma mesma região hidrológica são representadas por um reservatório equivalente de energia (CEPEL, 1999; SOUZA, *et al.*, 2012). Dessa forma, considera-se o SIN constituído por quatro subsistemas: Sudeste/Centro Oeste, Sul, Nordeste e Norte. As vazões mensais às usinas são também agregadas e tratadas como energias naturais afluentes através de um modelo do tipo PAR (SOUZA *et al.*, 2012).

Entende-se que a agregação em sistemas equivalentes é uma simplificação necessária a certos sistemas para tornar viável seu planejamento, contudo omite características individuais das usinas e peculiaridades hidrológicas das bacias nas quais elas operam. Como solução alternativa, propõe-se neste artigo a modelagem individual das vazões mensais afluentes às usinas do SIN através da formulação CARMA, ou autorregressiva com médias móveis contemporânea (SALAS *et al.*, 1980; CAMACHO *et al.*, 1987b; HIPEL & MCLEOD, 1994). Para manter a parcimônia do modelo, a sazonalidade e estacionariedade estatística das séries são tratadas à parte.

Adicionalmente, propõe-se um método para amostragem de séries hidrológicas sintéticas geradas através do modelo estocástico, na intenção de reduzir o número total de cenários em estudos que exijam grande processamento computacional. Esta amostragem é realizada em duas etapas: (i) classificação das matrizes de séries sintéticas de acordo com um critério de similaridade com a matriz de séries históricas e (ii) seleção das séries de forma que a amostra final contenha elementos de diferentes probabilidades de ocorrência na população, garantindo assim representatividade de eventos hidrológicos diversos.

O restante do artigo está organizado da seguinte forma: o próximo item e seus subitens descrevem o modelo utilizado para gerar as séries, bem como a técnica de amostragem proposta. Na sequência é apresentada uma descrição sucinta do sistema utilizado e suas características hidrológicas, seguido dos resultados e da conclusão do artigo.

## MATERIAIS E MÉTODOS

### Descrição do modelo de geração

A geração dos cenários sintéticos mensais de vazão se deu através de um modelo estocástico linear, não periódico, multivariado do tipo CARMA, ou autorregressivo com médias móveis contemporâneo. Sejam os vetores  $\mathbf{Z}_t = (Z_{t1}, Z_{t2}, \dots, Z_{tk})'$  e  $\mathbf{a}_t = (a_{t1}, a_{t2}, \dots, a_{tk})'$ , definidos para  $k$  séries temporais no tempo  $t$ . O modelo CARMA(p,q) é dado, genericamente, pela equação (1):

$$\boldsymbol{\varphi}_i(B)\mathbf{Z}_{ti} = \boldsymbol{\theta}_i(B)\mathbf{a}_{ti}, \quad i = 1, 2, \dots, k \quad (1)$$

onde  $\boldsymbol{\varphi}_i$  é o  $i$ -ésimo operador AR de ordem  $p$  ( $p = \max(p_1, \dots, p_k)$ ):

$$\boldsymbol{\varphi}_i(B) = 1 - \boldsymbol{\varphi}_{i1}B - \boldsymbol{\varphi}_{i2}B^2 - \dots - \boldsymbol{\varphi}_{ip}B^p$$

Da mesma forma,  $\boldsymbol{\theta}_i$  é o  $i$ -ésimo operador MA de ordem  $q$  ( $q = \max(q_1, \dots, q_k)$ ):

$$\boldsymbol{\theta}_i(B) = 1 - \boldsymbol{\theta}_{i1}B - \boldsymbol{\theta}_{i2}B^2 - \dots - \boldsymbol{\theta}_{iq}B^q$$

$B$  é o operador de defasagem do modelo. Para respeitar as condições de invertibilidade e estacionariedade da formulação (Box *et al.*, 1994),  $\boldsymbol{\varphi}_i(B) = 0$  e  $\boldsymbol{\theta}_i(B) = 0$  devem permanecer fora do círculo unitário. Por fim, assume-se o vetor de resíduos independentes e normalmente distribuídos  $\mathbf{a}_t \sim NID(\mathbf{0}, \Delta)$ , sendo  $\Delta$  a matriz de variância-covariância de  $\mathbf{a}_t$ .

A principal diferença da formulação CARMA para um modelo ARMA multivariado tradicional reside no fato de que as matrizes de parâmetros do modelo contemporâneo são diagonais. Assim, ele respeita, além das estatísticas básicas, as autocorrelações individuais de cada série histórica. A correlação espacial é preservada sucintamente (*lag zero*) através do vetor de resíduos, modelado a partir da equação (2):

$$\mathbf{a}_t = \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_t \quad (2)$$

onde  $\boldsymbol{\varepsilon}_t$  é um vetor de variáveis aleatórias independentes e identicamente distribuídas  $\boldsymbol{\varepsilon}_t \sim IID(\mathbf{0}, \mathbf{I})$  e  $\boldsymbol{\Lambda}$  é uma matriz-parâmetro de tamanho  $k \times k$ .

Camacho *et al.* (1987a) oferecem detalhamento teórico do modelo CARMA para o caso bivariado. Os mesmos autores estendem o estudo para o caso geral CARMA(p,q), descrevendo de forma minuciosa suas propriedades estatísticas (CAMACHO *et al.*, 1987b). O modelo contemporâneo é

tratado também em Hipel & McLeod (1994) e Salas *et al.* (1980). Aplicações práticas em hidrologia podem ser conferidas nos trabalhos de Haltiner & Salas (1988), Rasmussen *et al.* (1996), Wang (2008) e Stedinger *et al.* (1985). Neste último, em particular, os autores comparam o modelo CARMA com uma tradicional formulação multivariada ARMA, ambos de primeira ordem, concluindo que a performance dos dois não apresentou diferenças significativas.

Para o ajuste do modelo, aplicou-se transformação logarítmica às séries de vazões, suficientes para estabilizar suas variâncias e aproximá-las de uma distribuição gaussiana (KELMAN *et al.*, 1987). Ademais, a sazonalidade das séries foi removida através de padronização individual por média e desvio padrão (HIPEL & MCLEOD, 1994). Como mencionado anteriormente, esse procedimento foi empregado em detrimento do uso de um modelo periódico na intenção de manter o modelo final parcimonioso.

**Identificação do modelo**

A técnica mais tradicional de se identificar um modelo estocástico linear é através da comparação gráfica entre as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP). É sabido, entretanto, que esse método pode se tornar subjetivo quando a série apresenta indícios das porções AR e MA conjuntas em sua estrutura. Ademais, Souza *et al.* (2012) mostram que imprecisões na determinação de intervalos de confiança das FAC e FACP podem levar à superestimação da ordem do modelo.

Para contornar esses problemas, podem ser empregados os chamados Critérios de Informação, que são funções matemáticas estruturadas de forma a confrontar o ajuste oferecido pelo modelo com o número de parâmetros utilizado por ele. Os Critérios de Informação de Akaike (AIC – AKAIKE, 1974) e Bayesiano (BIC – SCHWARTZ, 1978) são os mais difundidos na literatura especializada. Na comparação entre eles, estudos mostraram que o BIC fornece estimadores consistentes, em detrimento do AIC (HANNAN, 1980; BERAN *et al.*, 1998). Outros autores ressaltam, ainda, que o AIC tende a superestimar a ordem de modelos predominantemente autorregressivos (SHITTU & ASEMOTA, 2009). Por esse motivo, adotou-se o Critério BIC como padrão para a decisão. Matematicamente, o BIC é definido através da equação (3):

$$BIC = -2 \ln L(r) + \ln nr \tag{3}$$

onde  $L(r)$  é a função de verossimilhança para  $r$  parâmetros e  $n$  é o número de elementos da série. A equação (3) é aplicada para todos os modelos candidatos e a formulação escolhida é a que apresentar o mínimo BIC. Para o presente trabalho, cinco modelos foram testados para cada série: CAR(1), CAR(2), CARMA(1,1), CARMA(2,1) e CARMA(2,2). A restrição de ordem máxima dois se deu por ela ser suficiente para o ajuste de modelos lineares a séries estacionárias (BOX *et al.*, 1994).

**Estimação dos parâmetros do modelo**

Ao utilizar matrizes diagonais de parâmetros para as porções AR e MA, o modelo contemporâneo permite que sejam estimados parâmetros de forma isolada para cada série. Dessa maneira, aplicou-se o método da máxima verossimilhança de forma individual a cada usina, na forma mostrada pelas equações (4):

$$\max \left[ \ln L(\varphi, \theta) = -n \ln \sigma_a^2 - \frac{S(\varphi, \theta)}{2\sigma_a^2} \right] \tag{4}$$

s.a.

$$-1 \leq \varphi \leq 1 \tag{4a}$$

$$-1 \leq \theta \leq 1 \tag{4b}$$

onde  $\sigma_a^2$  é a variância da série de resíduos e  $S(\varphi, \theta)$  é a função soma dos quadrados dos resíduos. A equação (4) é a função de log-verossimilhança tal qual definida em Box *et al.* (1994). As restrições (4a) e (4b) são impostas aos parâmetros para que sejam respeitadas as condições de estacionariedade de invertibilidade do modelo. Aplicou-se otimização através de Pontos Interiores (BYRD *et al.*, 2000) utilizando estimativas amostrais como valores iniciais para  $\varphi$  e  $\theta$ .

A geração do campo espacial correlacionado é feita através da matriz-parâmetro  $\Lambda$ , mostrada na equação (2). Para tanto, pode-se estimá-la buscando solução para a equação (5):

$$\Lambda \Lambda' = M \tag{5}$$

De acordo com Salas *et al.* (1980) qualquer solução que satisfaça a equação (5) é uma solução válida. No presente estudo, primeiramente estimou-se a matriz  $M$  substituindo-a pela matriz de correlações cruzadas dos resíduos (MATALAS, 1967). Na sequência, aplicou-se a decomposição de Cholesky sobre  $M$ , resultando na matriz triangular  $\Lambda$ . Ressalta-

se que este procedimento é válido somente para matrizes  $\mathbf{M}$  positivo-definidas. Caso essa condição não seja atendida, recomenda-se consulta ao estudo de Kuczera (1987).

### Validação do modelo

Seguindo recomendações de Haltiner & Salas (1988), o modelo foi validado em três etapas: (i) validação teórica, (ii) validação de estatísticas de curto termo e (iii) validação de estatísticas de longo termo. A primeira etapa se refere às propriedades dos resíduos obtidos com o modelo ajustado. Foram feitas inferências acerca dos pressupostos adotados para os resíduos na teoria de Box & Jenkins: independência temporal e espacial, homocedasticidade e distribuição aproximadamente normal.

A segunda e terceira etapas tratam de verificações comparativas entre as séries históricas e sintéticas geradas. Na segunda, a intenção é somente validar a implementação computacional do método, pois estes valores foram utilizados na construção e estimação do modelo. Determinaram-se médias, variâncias, coeficientes de assimetria, autocorrelações de lags 1 e 2 e a matriz de correlações espaciais.

Para a terceira etapa, os indicadores escolhidos estão fortemente relacionados com as épocas de estiagens e acumulação de água em reservatórios. Para avaliar os períodos secos, a análise se baseou na teoria das corridas (Haltiner & Salas, 1988), na qual são contados quantos elementos em sequência estão abaixo de um valor de corte. Cada conjunto de elementos com essa característica é chamado de corrida. Sendo  $R$  o total de corridas observado em uma determinada série, os seguintes dados (equações (6) a (9)) podem ser calculados:

$$\bar{l} = \frac{1}{R} \sum_{i=1}^R l_i \quad (6)$$

$$l_{max} = \max(l_1, l_2, \dots, l_R) \quad (7)$$

$$\bar{Q} = \frac{1}{R} \sum_{i=1}^M Q_i \quad (8)$$

$$Q_{max} = \max(Q_1, Q_2, \dots, Q_R) \quad (9)$$

onde  $\bar{l}$  é a média da duração das corridas,  $l_i$  é a duração da  $i$ -ésima corrida da série,  $l_{max}$  é a máxima duração de uma corrida,  $\bar{Q}$  é a afluência média por corrida e  $Q_{max}$  é a afluência máxima acumulada de

uma corrida. Considerou-se a média de longo termo de cada usina como valor de corte.

No caso dos períodos de cheia, utilizou-se o conceito de máximo déficit acumulado para um dado nível de regularização  $\delta$ . A determinação dos déficits é feita de acordo com a equação (10):

$$D_t = \max \left\{ \begin{array}{l} 0 \\ D_{t-1} - Z_t + \delta \bar{Z} \end{array} \right. \quad (10)$$

onde  $D_t$  é o déficit acumulado no tempo  $t$  ( $t = 1, 2, \dots, n$ ),  $Z_t$  é a afluência no tempo  $t$  e  $\bar{Z}$  é a média de longo termo da série. O máximo déficit acumulado é calculado através de  $D_{max} = \max(D_t)$ . O nível de regularização utilizado foi de 80%.

### Amostragem das séries sintéticas

A utilização de técnicas de amostragem estatística encontra aplicação em estudos nos quais se necessitam informações de uma população, mas não se dispõe de recursos suficientes para realizar inferências sobre todos os dados. No contexto do presente trabalho, os recursos são associados ao tempo computacional gasto por modelos que farão uso das séries sintéticas.

Para a diminuição do número dos cenários sintéticos gerados, duas considerações foram feitas: (i) o modelo de geração é multivariado, fazendo com que a amostragem individual prejudique a correlação espacial entre as usinas e (ii) entende-se que existem diversos eventos hidrológicos, com diferentes probabilidades de ocorrência, estimulando o emprego de técnicas de amostragem não equiprovável.

Dessa maneira, optou-se, numa primeira etapa, por trabalhar com as similaridades entre as matrizes de séries sintéticas e a matriz de séries históricas a partir da determinação de distâncias de Mahalanobis (DE MAESSCHALCK *et al.*, 2000). Sua formulação é mostrada na equação (11):

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\bar{\mathbf{x}} - \bar{\mathbf{y}})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \bar{\mathbf{y}})} \quad (11)$$

onde  $\boldsymbol{\Sigma}^{-1}$  é a inversa da matriz de covariâncias,  $\bar{\mathbf{x}}$  e  $\bar{\mathbf{y}}$  são os vetores de médias das séries histórica e sintética, respectivamente. Por considerar a matriz de covariâncias, a distância de Mahalanobis garante que as relações entre usinas sejam preservadas.

Em hidrologia, poucos foram os estudos que fizeram uso dessa métrica, podendo-se citar o trabalho de Corduas (2011), no qual foi utilizada a distância de Mahalanobis entre os coeficientes harmô-

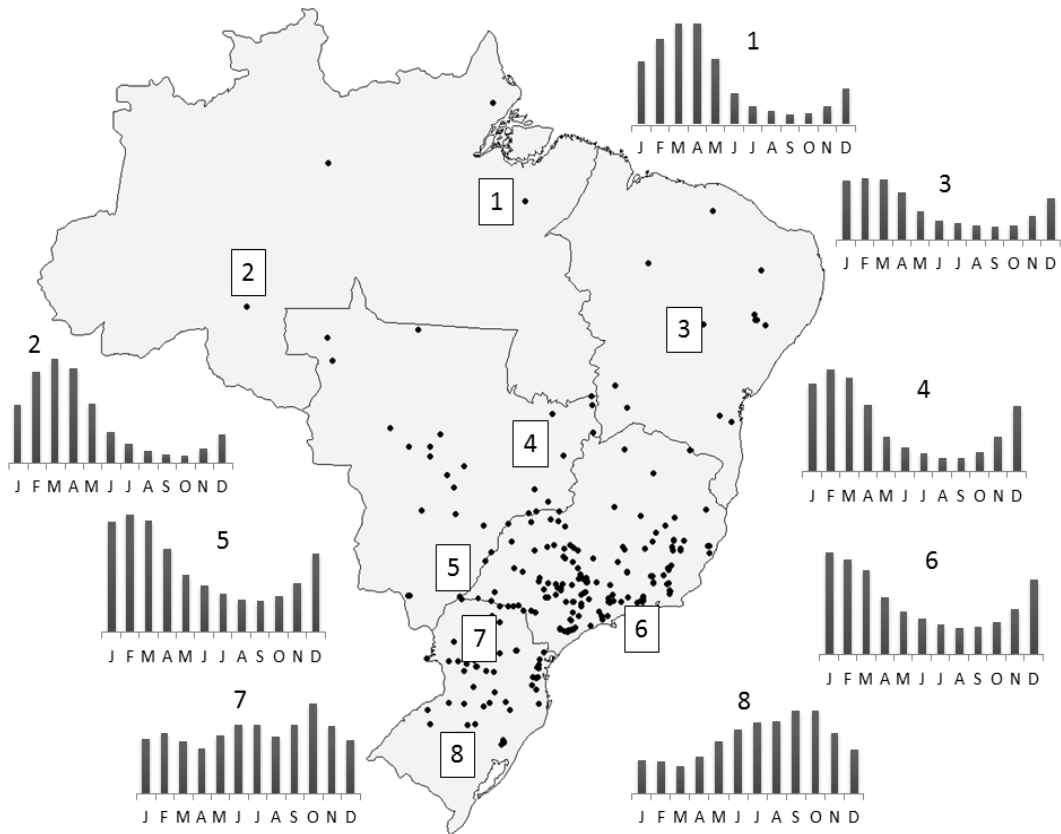


Figura 1 - Usinas utilizadas no estudo e histogramas de vazões médias mensais para algumas regiões

nicos da regressão que definiu a sazonalidade de séries de vazões de rios americanos, na intenção de classificá-las. No presente estudo, a lógica envolvida no método foi obter um ranque das séries sintéticas de acordo com sua semelhança com a série histórica. Quanto menor o valor da distância de Mahalanobis, mais próxima da série histórica é a série sintética.

Na sequência, aplicou-se a técnica de amostragem estratificada tradicional, adotando cinco estratos como padrão. Para determinar o tamanho global da amostra (equivalente à soma do número de elementos em cada estrato), aplicou-se a equação (12), deduzida para calcular o tamanho da amostra na intenção de estimar a média de uma população finita (COCHRAN, 1977):

$$n = \frac{Z_{\alpha}^2 \sigma^2 N}{d^2(N-1) + Z_{\alpha}^2 \sigma^2} \quad (12)$$

onde  $Z_{\alpha}$  é a abscissa normal padrão para nível de significância  $\alpha$ ,  $\sigma^2$  é a variância populacional,  $N$  é o tamanho da população (número de séries geradas)

e  $d$  é o erro amostral esperado. Adotou-se como nível de significância 95%, enquanto que para o erro amostral foi considerado 2% do valor médio populacional, equivalente à média das médias das séries sintéticas de cada usina.

## DESCRIÇÃO DA ÁREA DE ESTUDO

O modelo CARMA foi aplicado a um conjunto de 146 séries de aflúncias às usinas hidrelétricas pertinentes ao SIN (referência: dez./2007). A Figura 1 exhibe a distribuição das usinas sobre o território brasileiro, juntamente com os histogramas de vazões médias mensais para oito regiões distintas. As divisões no mapa se referem aos subsistemas que constituem o SIN: as regiões 1 e 2 se localizam no subsistema Norte, a região 3 no subsistema Nordeste, as áreas 4 a 6 representam o subsistema Sudeste/Centro-Oeste e as regiões 7 e 8 se localizam no subsistema Sul.

Uma simples análise nos histogramas permite notar a diversidade do comportamento hidrológico existente do território brasileiro. Séries com sazonalidade bem definida no norte do país contrastam com distribuições mais regulares na região sul. As variações são percebidas também em regiões localizadas dentro de um mesmo subsistema, como no caso do primeiro quarto do ano nas áreas 4 a 6. Ao utilizar técnicas de geração que consideram agregação de reservatórios, por exemplo, essas diferenças são ignoradas. É justamente nesse ponto que a modelagem individual se torna atrativa.

As séries históricas de vazões mensais foram coletadas do Operador Nacional do Sistema Elétrico (ONS), limitando-se ao período entre jan./1931 e dez./2007 (total de 924 meses) para todas as usinas. Os registros se referem às afluições naturalizadas e se encontram todos consistidos e sem falhas. A única análise realizada sobre as séries foi no tocante à sua condição de estacionariedade estatística. Em concordância com o propósito de utilizar um modelo parcimonioso para a geração dos cenários, optou-se por não utilizar formulações com séries diferenciadas (modelos ARIMA), por acarretar a necessidade de estimação de uma carga extra de parâmetros.

Todo o conjunto de séries foi submetido a cinco testes estatísticos (t-Student, Cox-Stuart, Wilcoxon, Spearman e Mann-Kendall) para a detecção de variações na média e tendências. Como resultado, 75 das 146 séries apresentaram indícios de não estacionariedade e foram submetidas a um procedimento de correção. Os detalhes sobre a execução dos testes e correção das séries não estacionárias podem ser conferidos no estudo de Detzel *et al.* (2011).

## RESULTADOS E DISCUSSÕES

### Resultados do modelo CARMA

Para validar o método utilizado, foram geradas 2000 séries sintéticas para cada usina, todas com o mesmo comprimento da série histórica. A primeira fase de verificações foi referente aos resíduos do modelo, obtendo-se resultados positivos para a grande maioria dos testes.

A Tabela 1 exibe resultados comparativos das estatísticas de curto termo entre as séries históricas e as médias das séries sintéticas para oito usinas do estudo. Os números na primeira coluna coincidem com as regiões apontadas na Figura 1. Nota-se que, de um modo geral, as estatísticas foram muito bem reproduzidas pelo modelo estocástico. Imprecisões

são observadas justamente nas usinas que operam em regiões nas quais a hidrologia não possui épocas sazonais bem definidas.

A fim de se investigar o diferente desempenho do modelo nestas regiões de hidrologia distinta, montaram-se os gráficos da Figura 2. Devido à limitação de espaço, foram detalhadas as médias mensais somente para as usinas de Tucuruí, na região Norte e Machadinho, na região Sul.

Percebe-se que, apesar de ter as médias mensais bem reproduzidas, a usina de Machadinho apresentou alguns desvios sutis, não observados na usina de Tucuruí. Os resultados para as demais usinas do estudo foram semelhantes nesse aspecto, mostrando-se mais precisos em usinas que operam em regiões com sazonalidade anual bem definida.

A Figura 3 representa de forma esquemática as matrizes de correlações espaciais históricas e sintéticas geradas pelo modelo contemporâneo. Adotou-se uma escala cinza, na qual quanto mais escuro mais o coeficiente se aproxima da correlação perfeita. A escala tem seu mínimo em -0,40, correlação histórica mais baixa encontrada entre as usinas do estudo.

Nota-se que, mesmo com o tamanho e complexidade do sistema modelado, as correlações especiais foram respeitadas apropriadamente. As pequenas “ilhas” de correlações elevadas, se referem às usinas que operam em cascata num mesmo rio. Por outro lado, a grande faixa de correlações negativas aponta o comportamento hidrológico contrário entre usinas localizadas no subsistema Sul e as usinas que operam nos demais subsistemas. Para um sistema interligado como o SIN, esse comportamento é positivo, visto que épocas de seca com baixa capacidade de geração hídrica em um subsistema podem ser complementadas pela geração de usinas localizadas em outro subsistema. Por esse motivo, a reprodução desta característica pelo modelo estocástico é de grande importância.

A análise das estatísticas de longo termo complementa a validação do modelo aqui utilizado. A Tabela 2 mostra estes índices para as mesmas oito usinas selecionadas anteriormente. Da mesma forma que as estatísticas de curto termo, os resultados foram bem reproduzidos pelo modelo, exceto pelo máximo déficit acumulado que apresentou alguns resultados regulares, principalmente para usinas com afluições de grande volume. Ainda assim, a análise dessas estatísticas permite diferenciar o comportamento hidrológico das usinas: no caso das corridas, por exemplo, a duração média é semelhante em todas as usinas, contudo a duração máxima é muito maior para usinas do Sul.

Tabela 1 – Estatísticas de curto termo para oito usinas

Região – Usina	Tipo	MLT (m <sup>3</sup> /s)	D. Padrão (m <sup>3</sup> /s)	Assim.	Autocorr. lag1	Autocorr. lag2	Mínimo (m <sup>3</sup> /s)	Máximo (m <sup>3</sup> /s)
<b>1 – Tucuruí</b>	SH	11.003	9.241	1,10	0,81	0,44	1.269	51.539
[Modelo]	SS	11.019	9.276	1,13	0,82	0,45	1.200	51.891
<b>2 – Samuel</b>	SH	352	293	0,79	0,80	0,41	16	1.339
[AR(1)]	SS	352	294	0,78	0,81	0,41	20	1.299
<b>3 – Sobradinho</b>	SH	2.687	1.976	1,64	0,76	0,41	506	15.676
[AR(1)]	SS	2.693	1.998	1,68	0,79	0,46	484	14.518
<b>4 - Serra da Meas</b>	SH	778	697	2,14	0,69	0,36	97	6.163
[ARMA(1,1)]	SS	777	689	2,11	0,75	0,43	81	5.614
<b>5 - Ilha Solteira</b>	SH	5.588	3.278	1,23	0,77	0,43	1.387	20.314
[ARMA(2,1)]	SS	5.596	3.345	1,50	0,78	0,45	1.405	23.725
<b>6 – Camargos</b>	SH	132	84	1,63	0,70	0,39	34	576
[ARMA(2,1)]	SS	132	83	1,72	0,74	0,42	29	622
<b>7 - Salto Osório</b>	SH	1.188	888	2,39	0,49	0,20	126	8.473
[AR(1)]	SS	1.191	926	3,00	0,50	0,24	101	9.504
<b>8 – Machadinho</b>	SH	814	654	2,02	0,48	0,22	57	5.925
[AR(2)]	SS	826	749	3,28	0,48	0,24	48	7.888

SH – Série Histórica; SS – Série Sintética.

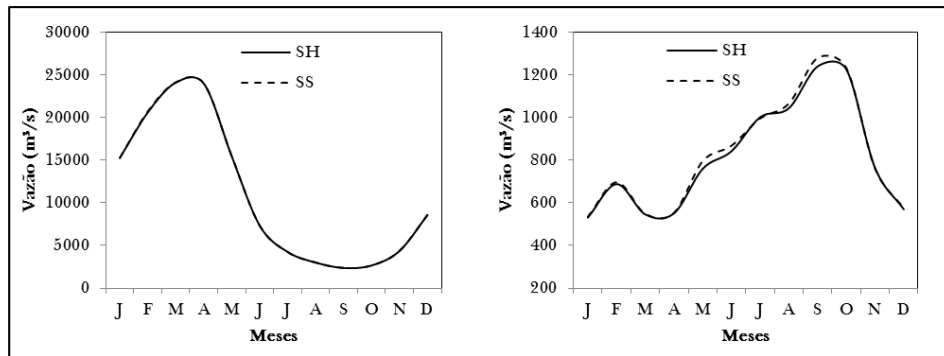


Figura 2 - Vazões históricas (SH) e sintéticas (SS) médias mensais para Tucuruí (esquerda) e Machadinho (direita)

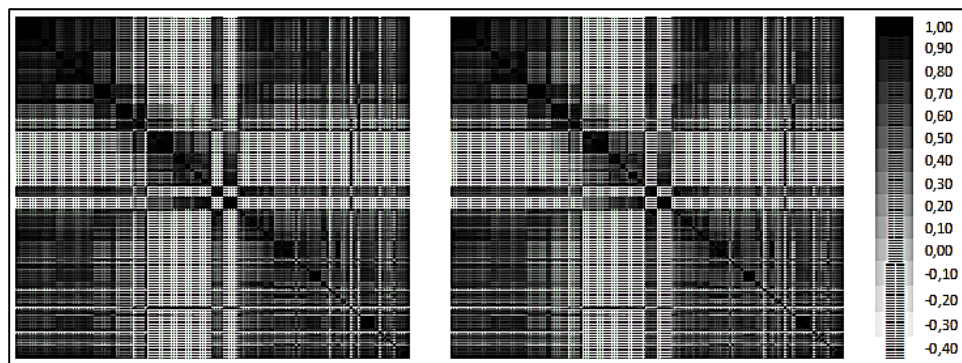


Figura 3 - Matrizes de correlações espaciais históricas (esquerda) e sintéticas (direita)

**Tabela 2 – Estatísticas de longo termo para oito usinas**

Região – Usina [Modelo]	Tipo	Corridas				Déficits Acumulados		
		Nº	Duração Média	Duração Máxima	Vazão Média (m³/s)	Vazão Máxima (m³/s)	Máximo (m³/s)	Médio (m³/s)
1 – Tucuruí	SH	77	7	10	34.113	50.593	155.984	23.243
[AR(1)]	SS	77	7	11	33.168	56.063	88.401	19.247
2 – Samuel	SH	77	7	8	963	1.350	1.642	454
[AR(1)]	SS	77	7	9	957	1.520	1.659	459
3 – Sobradinho	SH	79	7	11	10.433	16.040	30.631	4.554
[AR(1)]	SS	76	8	19	10.849	26.733	23.939	4.364
4 - Serra da Meas	SH	79	7	18	2.720	7.150	10.565	1.608
[ARMA(1,1)]	SS	76	8	20	2.860	7.487	8.857	1.772
5 - Ilha Solteira	SH	75	7	20	25.768	55.294	39.572	5.221
[ARMA(2,1)]	SS	76	7	18	25.944	61.580	32.745	5.058
6 – Camargos	SH	76	8	20	613	1.313	820	125
[ARMA(2,1)]	SS	76	8	19	611	1.473	936	146
7 - Salto Osório	SH	99	5	21	3.554	15.269	10.764	1.976
[AR(1)]	SS	97	6	24	3.931	15.141	13.182	2.379
8 – Machadinho	SH	94	6	27	2.433	7.964	10.432	2.075
[AR(2)]	SS	96	6	23	2.563	9.422	10.585	2.064

SH – Série Histórica; SS – Série Sintética.

**Amostragem das séries sintéticas**

Na aplicação da técnica de amostragem proposta, foram encontrados problemas relativos ao alto grau de colinearidade nos dados. Esta ocorrência impossibilita a determinação da inversa da matriz de covariâncias e, conseqüentemente, das distâncias de Mahalanobis. O problema da colinearidade na métrica de Mahalanobis não é novo, sendo reportado por Jouan-Rimbaud *et al.* (1998) e De Maesschalek *et al.* (2000), que ligam essa condição à inconsistências presentes na obtenção dos dados. Para os registros de vazão aqui utilizados, a colinearidade é consequência dos métodos de regressão aplicados entre os diferentes postos de medição, na intenção de corrigir falhas ou estender séries de menor duração. Para contornar este impasse, a recomendação é reduzir o tamanho do problema até que seja possível a determinação da inversa da matriz de covariâncias. No presente estudo, o critério utilizado foi selecionar as usinas cuja capacidade instalada fosse superior a 1 GW e todas as usinas localizadas nas cabeceiras de suas respectivas bacias hidrográficas. Dessa forma, o sistema utilizado para amostragem reduziu de 146 para 62 usinas.

Na primeira etapa do estudo, as distâncias de Mahalanobis foram determinadas entre a série histórica e cada matriz de séries sintéticas, resultando em 2000 distâncias. Na sequência, calculou-se o tamanho global da amostra, chegando-se ao valor médio de n=300 séries sintéticas por usina.

A amostragem estratificada foi feita sobre o total de N=2000 distâncias, distribuindo-as em cinco estratos  $N_h$  ( $h = 1, \dots, 5$ ). Deste conjunto foram determinados os pesos dos estratos  $W_h = N_h/N$ , a partir dos quais a amostra de n=300 foi obtida. Os parâmetros da amostragem são expressos na Tabela 3.

**Tabela 3 – Parâmetros da amostragem estratificada**

#	$N_h$	$W_h$	$n_h$	$s_h^2$
1	118	0,06	18	0,08
2	368	0,19	55	0,04
3	682	0,34	102	0,07
4	566	0,28	85	0,10
5	266	0,13	40	0,42
<b>Total</b>	<b>2000</b>	<b>1,00</b>	<b>300</b>	<b>3,8x10<sup>-4</sup></b>



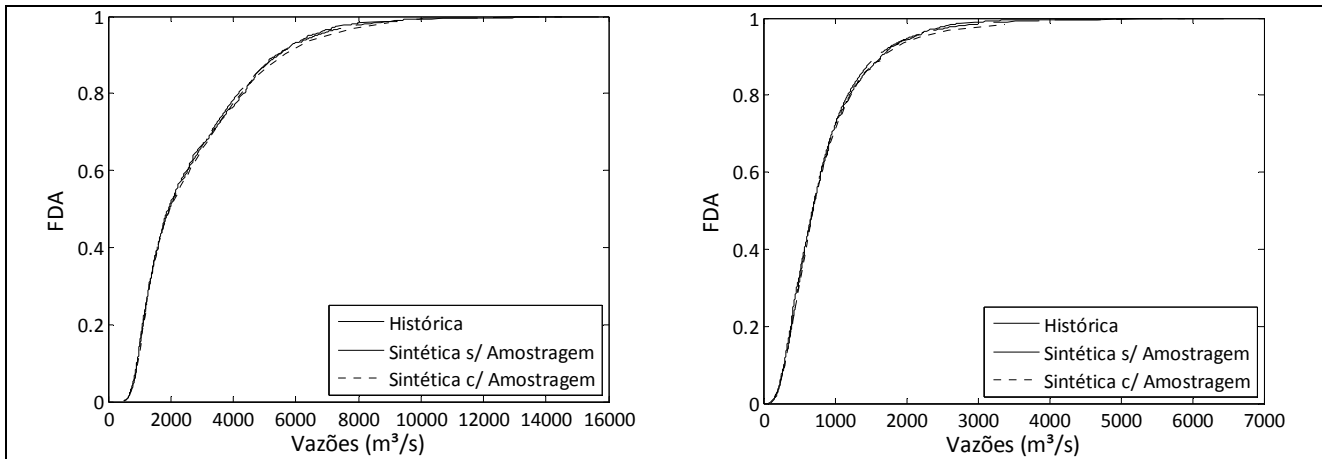


Figura 4 - Distribuições de probabilidades empíricas acumuladas para Sobradinho (esquerda) e Salto Osório (direita)

Analisando as variâncias individuais  $s_h^2$ , nota-se que o quinto estrato apresentou um valor discrepante em relação aos demais. Isso ocorreu por ser neste estrato que se encontram as distâncias com maior valor absoluto. Em outras palavras, o estrato cinco reuniu o conjunto de séries mais dissimilares em relação à série histórica. Ainda assim, a variância total da amostra estratificada resultou em um valor muito baixo, validando o método de estratificação utilizado.

Uma vez compostas as amostras, elas foram submetidas às análises de distribuição empírica de probabilidades. A Figura 4 mostra a comparação entre as funções de distribuição de probabilidades empíricas acumuladas (FDA) da série histórica, séries sintéticas sem amostragem ( $N=2000$ ) e séries sintéticas com amostragem ( $n=300$ ) para duas usinas: Sobradinho e Salto Osório. Percebe-se que a amostragem não afetou significativamente as funções empíricas em nenhum dos casos. As maiores diferenças foram observadas na porção superior, relativas ao intervalo entre 0,8 e 1,0, no qual as séries amostradas ligeiramente subestimaram as vazões de maior intensidade. Estes resultados foram também verificados para as demais usinas do estudo.

Desta maneira, mostra-se ser possível a redução do número de cenários sintéticos sem perdas significativas na representatividade hidrológica local, alternativa que pode se mostrar interessante em estudos relativos a sistemas complexos e de grande porte. Adicionalmente, esse método torna possível a seleção de séries sintéticas com características particulares de períodos de estiagem ou cheias para serem usados em estudos específicos.

## CONCLUSÕES

A modelagem de grandes sistemas mostra-se desafiadora à medida que se aumentam a complexidade e o número de variáveis envolvidas. O caminho comumente encontrado é simplificá-las para tornar viável a solução do problema. No entanto, estas simplificações podem omitir características importantes do sistema, colocando em risco a confiabilidade dos resultados.

No caso de grandes sistemas hídricos, a agregação de reservatórios é uma solução empregada frequentemente, principalmente em análises estocásticas. No presente artigo, apresentou-se um método alternativo para modelagem de séries de vazões, baseado em formulação autorregressiva contemporânea. Essa classe de modelos foi elaborada com princípios parcimoniosos, tornando-se atrativa para casos de maior complexidade. Ainda assim, a aplicação destes modelos em hidrologia é limitada.

Neste trabalho, as vazões de 146 usinas hidrelétricas integrantes do SIN foram modeladas com sucesso pela formulação CARMA. O modelo foi capaz de preservar estatísticas de curto e longo termos, além de reproduzir com precisão a matriz de correlações espaciais.

A modelagem individual de usinas permite a incorporação de elementos específicos, como desvios, bombeamentos e transposição de rios, comuns a diversos sistemas hídricos. Outra grande vantagem é a possibilidade de deplecionamento diferencial entre reservatórios, consideração impossibilitada quando se trabalha com sistemas agregados.

Além da modelagem mostrada, propôs-se a redução do número de séries sintéticas geradas pelo modelo estocástico aplicando-se uma técnica híbrida de classificação de séries e amostragem não equiprovável. Mostrou-se, a partir de distribuições empíricas de probabilidades, que a técnica não distorce as características obtidas com o modelo estocástico. No entanto, problemas com a colinearidade dos dados forçaram a redução do sistema para tornar possível a classificação das séries. Dessa forma, sugere-se o refinamento dessa técnica como assunto a ser abordado em estudos futuros.

## AGRADECIMENTOS

Esta pesquisa/trabalho foi possível graças ao financiamento da ANEEL através do Projeto Estratégico de Pesquisa e Desenvolvimento – ANEEL PE-6491-0108/2009, “Otimização do Despacho Hidrotérmico”, com o apoio das seguintes concessionárias: COPEL, DUKE, CGTF, CDSA, BAESA, ENERCAN, CPFL PAULISTA, CPFL, PIRATININGA, RGE, AES TIETÊ, AES URUGUAIANA, ELETROPAULO, CEMIG e CESP. Os autores agradecem as contribuições dos revisores anônimos.

## REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, N. 6, p. 716-723, 1974.

ARVANITIDIS, N. V., ROSING, J. Composite representation of a multireservoir hydroelectric power system. *IEEE Transactions on Power Apparatus and Systems*, Vol. 2, p. 319-326, 1970.

BERAN, J., BHANSALI, R. J., OCKER, D. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*. Vol. 85, N. 4, p. 921-934, 1998.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. *Time Series Analysis: Forecasting and Control*, 3 ed. Prentice Hall, New Jersey, 1994.

BYRD, R.H., GILBERT, J. C., NOCEDAL, J. A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming. *Mathematical Pro-*

*gramming*, Vol. 89, N. 1, p. 149–185, 2000.

CAMACHO, F., McLEOD, A. I., HIPEL, K. W. Contemporaneous bivariate time series. *Biometrika*, Vol. 74, N. 1, p. 103-113, 1987a.

CAMACHO, F., McLEOD, A. I., HIPEL, K. W. Multivariate contemporaneous ARMA model with hydrological applications. *Stochastic Hydrology and Hydraulics*, Vol. 1, p. 141-154, 1987b.

CAN, İ., TOSUNOĞLU, F., KAHYA, E. Daily streamflow modelling using autoregressive moving average and artificial neural networks models: case study of Çoruh basin, Turkey. *Water and Environment Journal*, Vol. 26, N. 4, p. 567–576, 2012.

CASTELLANO-MÉNDEZ, M.; GONZÁLEZ-MANTEIGA, W.; FEBRERO-BANDE, M.; PRADA-SÁNCHEZ, J. M.; LOZANO-CALDERÓN, R. Modelling of the monthly and daily behaviour of the runoff of the Xalas river using Box-Jenkins and neural networks methods. *Journal of Hydrology*, Vol. 296, p. 38-58, 2004.

CEPEL – Centro de Pesquisas de Energia Elétrica, Manual de Referência do modelo NEWAVE, Relatório Técnico, 1999.

CORDUAS, M. Clustering streamflow time series for regional classification. *Journal of Hydrology*, Vol. 407, N. 1-4, p. 73-80, 2011.

COCHRAN, W.G. *Sampling Techniques*, 3 ed. John Wiley & Sons, Inc., New York, 1977.

CRUZ Jr., G., SOARES, S. Non-uniform Composite Representation of Hydroelectric Systems for Long-Term Hydrothermal Scheduling. *IEEE Transactions on Power Systems*. Vol 11, N. 2, p. 702-707, 1996.

DE MAESSCHALCK, R., JOUAN-RIMBAUD, D., MASSART, D. L. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, Vol. 50, N.1, p. 1-18, 2000.

DETZEL, D. H. M., BESSA, M. R., VALLEJOS, C. A. V., SANTOS, A. B., THOMSEN, L. S., MINE, M. R. M., BLOOT, M. L., ESTRÓCIO, J. P. Estacionariedade das Afluências às Usinas Hidrelétricas Brasileiras. *Revista Brasileira de Recursos Hídricos*, Vol. 16, N. 3, p. 95-111, 2011.

GUIMARÃES, R. C., SANTOS, E. G. Principles of

- Stochastic Generation of Hydrologic Time Series for Reservoir Planning and Design: Case Study. *Journal of Hydrologic Engineering*, Vol. 16, N. 11, p. 891–898, 2011.
- HALTINER, J. P., SALAS, J. D. Development and testing of a multivariate, seasonal ARMA(1,1) model. *Journal of Hydrology*, Vol. 104, p. 247-272, 1988.
- HANNAN, E. J. The estimation of the order of an ARMA process. *Annals of Statistics*, Vol. 8, N. 5, p. 1071-1081, 1980.
- HAO, Z., SINGH, V. P. Single-site monthly streamflow simulation using entropy theory. *Water Resources Research*, Vol. 47, N. W09528, 2011.
- HIPEL, K. W., McLEOD, A. I. Time series modelling of water resources and environmental systems, 1994. Disponível em: <http://www.stats.uwo.ca/faculty/aim/1994Book/>. Acesso em: 12/11/2012.
- JACKSON, B. B. The use of streamflow models in planning. *Water Resources Research*, Vol. 11, N. 1, p. 54-63, 1975.
- JOUAN-RIMBAUD, D., MASSART, D. L., SABY, C. A., PUEL, C. Determination of the representativity between two multidimensional data sets by a comparison of their structure. *Chemometrics and Intelligent Laboratory Systems*, Vol. 40, N. 2, p. 129-144, 1998.
- KELMAN, J. Modelos estocásticos no gerenciamento de recursos hídricos. In: Barth *et al.* (Eds.) Modelos para gerenciamento de recursos hídricos I. São Paulo: Nobel/ABRH, 1987.
- KUCZERA, G. On maximum likelihood estimators for the multisite lag-one streamflow model: complete and incomplete data cases. *Water Resources Research*, Vol. 23, N. 4, p. 641-645, 1987.
- LUNA, I., BALLINI, R., SOARES, S.; SILVA FILHO, D. Fuzzy inference systems for synthetic monthly inflow time series generation. In: Anais do EUSFLAT, p. 1060-1065, Aix-les-Bains, França, 2011.
- LUNGU, E. M.; SEFE, F. T. K. Stochastic analysis of monthly streamflows. *Journal of Hydrology*, Vol. 126, p. 171-182, 1991.
- MAHEEPALA, S., PERERA, B. J. C.. Monthly hydrologic data generation by disaggregation. *Journal of Hydrology*, Vol. 178, p. 277-291, 1996.
- MATALAS, N. C. Mathematical assessment of synthetic hydrology. *Water Resources Research*, Vol. 3, N. 4, p. 937-945, 1967.
- MEJIA, J. M., ROUSSELLE, J. Disaggregation models in hydrology revisited. *Water Resources Research*, Vol. 12, N. 2, p. 185-186, 1976.
- MONDAL, M. S., WASIMI, S. A. Generating and forecasting monthly flows of the Ganges river with PAR model. *Journal of Hydrology*, Vol. 323, p. 41-56, 2006.
- NIU, J., SIVAKUMAR, B. Scale-dependent synthetic streamflow generation using a continuous wavelet transform. *Journal of Hydrology*, Vol. 496, p. 71–78, 2013.
- PEREIRA, M. V. F., OLIVEIRA, G. C. , COSTA, C. C. G., KELMAN, J. Stochastic streamflow models for hydroelectric systems. *Water Resources Research*, Vol. 20, N. 3, p. 379-390, 1984.
- RASMUSSEN, P. F., SALAS, J. D., FAGHERAZZI, L., RASSAM, J.-C., BOBÉE, B. Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resources Research*, Vol. 32, N. 10, p. 3151-3160, 1996.
- SALAS, J. D., DELLEUR, J. W., YEVJEVICH, V., LANE, W. L. Applied modeling of hydrologic time series. Littleton, Colorado: Water Res Publications, 1980.
- SALAS, J. D., LEE, T. Nonparametric Simulation of Single-Site Seasonal Streamflows. *Journal of Hydrologic Engineering*, Vol. 15, N. 4, p. 284–296, 2010.
- SCHWARTZ, G. Estimating the dimension of a model. *Annals of Mathematical Statistics*, Vol. 6, N. 2, p. 461-464, 1978.
- ŞEN, Z. A mathematical model of monthly flow sequences. *Hydrological Sciences Journal*, Vol. 23, N. 2, p. 223-229, 1978.
- SHITTU, O. I., ASEMOTA, M. J. Comparison of criteria for estimating the order of autoregressive process: a Monte Carlo approach. *European Journal of Scientific Research*, Vol. 30, N. 3, p. 409-416, 2009.

SILVA, A. T., PORTELA, M. M. Stochastic Assessment of Reservoir Storage-Yield Relationships in Portugal. *Journal of Hydrologic Engineering*, Vol. 18, N. 5, p. 567-575, 2013.

SOUZA, R. C., MARCATO, A. L. M., DIAS, B. H., OLIVEIRA, F. L. C. Optimal operation of hydrothermal systems with hydrological scenario generation through bootstrap and periodic autoregressive models. *European Journal of Operational Research*, Vol. 222, p. 606-615, 2012

STEDINGER, J. R., LETTENMAIER, D. P., VOGEL, R. M. Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. *Water Resources Research*, Vol. 21, N. 4, p. 497-509, 1985.

THOMAS, H. A., FIERING, M. B. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: Maass, A. et al. (Eds.), *Design of Water Resources Systems*. Harvard Univ. Press, Cambridge, p. 459-493, 1962.

VALENCIA, D. R., SHAAKE Jr., J. C. Disaggregation processes in stochastic hydrology. *Water Resources Research*, Vol. 9, N. 3, p. 580-585, 1973.

WANG, Q. J. A Bayesian method for multi-site stochastic data generation: dealing with non-concurrent and missing data, variable transformation and parameter uncertainty. *Environmental Modelling & Software*, Vol. 23, p. 412-421, 2008.

WANG, W., HU, S., LI, Y. Wavelet Transform Method for Synthetic Generation of Daily Streamflow. *Water Resources Management*, Vol. 25, N. 1, p. 41-57, 2010.

### ***Synthetic Flows Scenarios for Large Water Systems Through Contemporary Models and Sampling***

#### ***ABSTRACT***

*Modeling large water systems proves to be challenging as the complexity and the number of variables involved increase. It is common to employ simplifications as, for example, aggregation of reservoirs that operate in the same region. However, these simplifications may omit important system features, making them inflexible and jeop-*

*ardizing the reliability of results. In this paper, an alternative method is presented for modeling streamflow series based on the CARMA formulation. The model is successfully applied to monthly streamflow from 146 hydroelectric plants that comprise the Brazilian Electric System, and is able to reproduce short and long term statistics and preserve the spatial correlations among the plants. Furthermore, a hybrid technique of classification and non-equiprobable stratification is proposed in order to reduce the number of synthetic series generated by the stochastic model. The intention is to make it possible to apply the generated series in studies requiring greater computer processing. It is shown that it is possible to reduce the number of scenarios without distorting the probability distribution obtained with the stochastic model.*

**Key-words:** *contemporary model, monthly synthetic series, large water systems, sampling.*