

USO DE SÉRIES SINTÉTICAS NO PLANEJAMENTO E OPERAÇÕES DE  
SISTEMAS HIDROTÉRMICOS

Jerson Kelman

CEPEL - Centro de Pesquisas de Energia Elétrica  
Caixa Postal 2754 - CEP 20001  
Cidade Universitária - Rio de Janeiro  
Brasil

RESUMO

Apresenta-se uma breve crítica do método da energia firme para avaliação do potencial energético de um sistema hidroelétrico. Como alternativa, sugere-se o uso de séries sintéticas. São revistos os princípios envolvidos na identificação do modelo estocástico e estimação dos parâmetros, bem como o procedimento para verificação e validação do modelo.

# USO DE SÉRIES SINTÉTICAS NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS HIDROTÉRMICOS\*

Jerson Kelman\*\*

## 1- INTRODUÇÃO

A produção energética de um sistema hidroelétrico depende, entre outras coisas, da série cronológica de vazões afluentes às diversas usinas que compõem o sistema. Na impossibilidade de se prever as afluências futuras, adota-se frequentemente a hipótese de que o registro de vazões observadas no passado, a chamada série histórica, será repetida.

Por exemplo, uma nova usina hidroelétrica é usualmente avaliada pela "energia firme" que ela acrescenta ao sistema. "Energia firme" é definida como a máxima demanda energética que o sistema pode atender, sem que ocorra alguma falha no suprimento, quando as vazões que afluem aos diversos aproveitamentos são idênticas àsquelas registradas na pior seca (período crítico) observada no passado.

A série histórica é apenas uma das possíveis realizações de um processo estocástico. Isto é, pode-se imaginar que a Natureza "sorteou" a série histórica, segundo algum conjunto de leis probabilísticas. Um novo sorteio redundaria numa outra série, diferente da histórica, mas igualmente provável. Um modelo estocástico aproxima este comportamento estocástico, de modo a permitir que se façam tantos sorteios artificiais quantos forem necessários para o estudo em foco.

---

\* Apresentado no Seminário Latinoamericano de Aproveitamento dos Recursos Hídricos, Universidad Nacional da Colombia, Medellin. (1987)

\*\* Pesquisador do CEPEL e Professor da COPPE-UFRJ, Rio de Janeiro, Brasil.

Cada sorteio estará associado a uma série de vazões, chamada de série sintética. Como estas séries serão todas distintas entre si, pode-se obter diversos resultados provenientes de simulações, ao invés de um único resultado que seria obtido caso apenas a série histórica estivesse disponível. Desta forma, a informação contida na série histórica pode ser mais completamente extraída, permitindo assim que se avalie riscos e incertezas pertinentes a um sistema hidroelétrico. Por exemplo, pode-se calcular para cada uma das séries sintéticas uma "energia firme sintética", à semelhança do que se faz com a série histórica. A probabilidade de não atendimento de uma demanda energética, dentro do horizonte de planejamento, pode ser estimada pela frequência com que as "energias firmes sintéticas" forem inferiores à esta demanda. A figura 1 ilustra o conceito para o caso em que a demanda é igual a energia firme histórica.

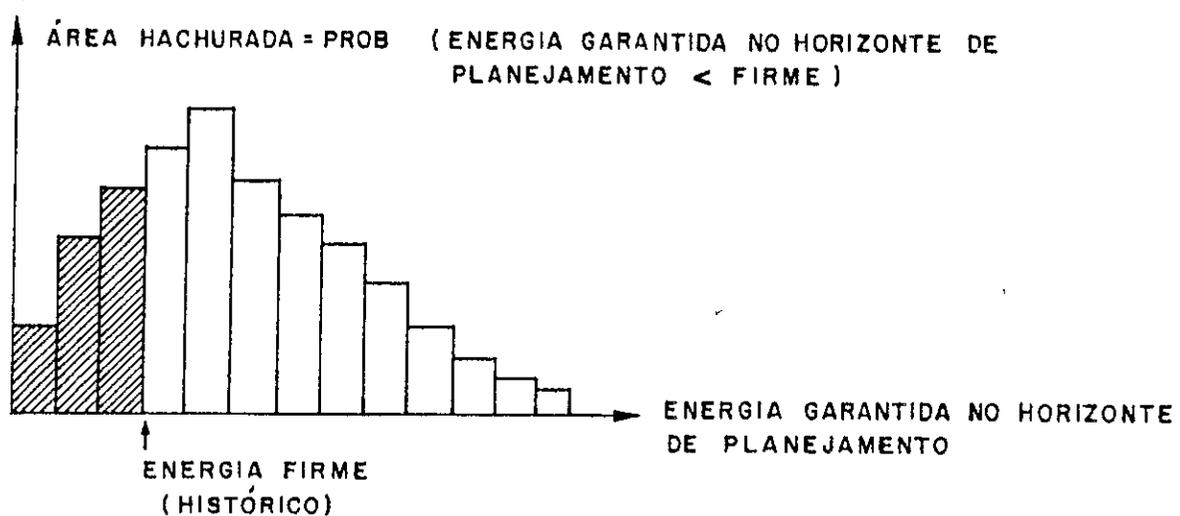


Figura 1 - Gráfico de frequência relativa de energia garantida no horizonte de planejamento

As séries sintéticas podem ser utilizadas para determinar a relação entre demanda energética e risco de não atendimento da demanda num ano qualquer. A figura 2 mostra uma curva típica que relaciona estas duas grandezas. O sistema hidroelétrico poderia ser considerado capaz de suprir a demanda  $e$ , que, como se vê, corresponde ao risco  $p$ , previamente selecionado.

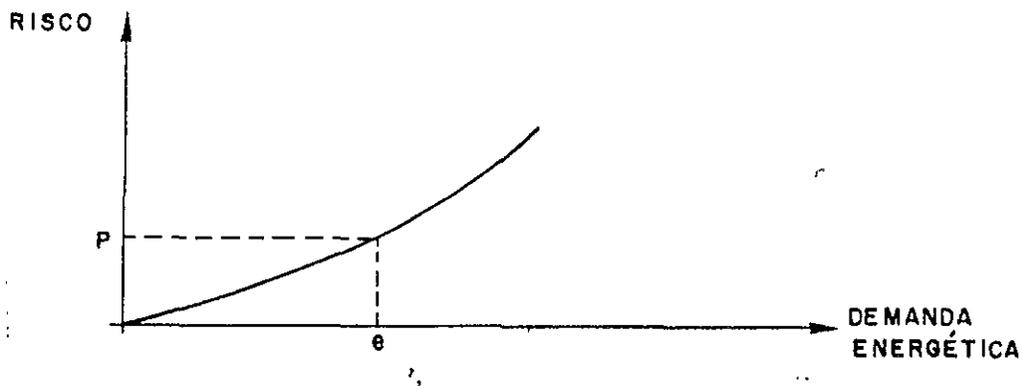


Figura 2 - Relação Demanda x Risco num ano qualquer

Na mesma linha de raciocínio, a contribuição de uma usina hidroelétrica para o sistema poderia ser considerada, por exemplo, como a média da produção de energia da usina ao longo de "períodos secos". Um período seco é o correspondente probabilístico do período crítico do histórico: um mês, na simulação com uma série sintética, estará contido num período seco se houver algum déficit no futuro e se o reservatório equivalente (Terry et al, 1986) não encher completamente durante o intervalo que vai deste mês até a ocorrência de déficit. Esta avaliação exigiria uma simulação com um modelo a usinas individualizadas que utilizasse a série sintética como dado de entrada. A simulação produziria também outros resultados de interesse. Por exemplo, a distribuição de probabilidades da potência disponível nas diversas usinas, devido à variações de nível nos reservatórios, como ilustrado na figura 3.

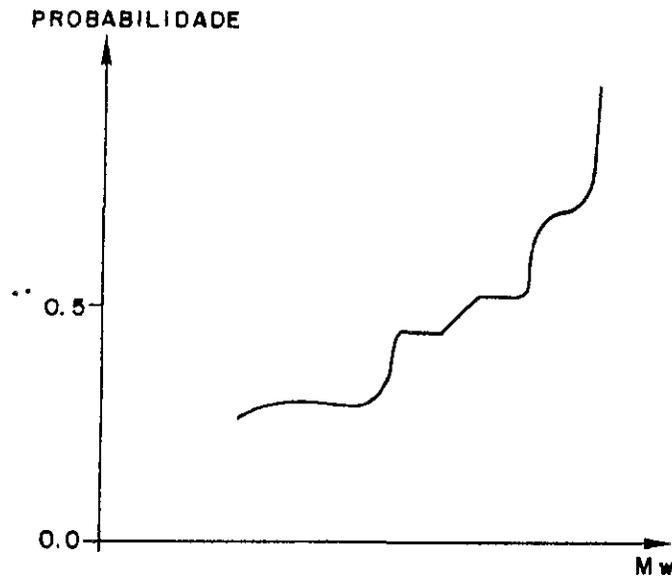


Figura 3 - Distribuição de potência disponível numa usina

Séries sintéticas podem ser utilizadas também para determinar a política de operação para o sistema através de regras que sejam em média melhores para um grande número de séries, e não uma regra feita sob medida para a série histórica, como às vezes se faz. Isto é, a alternativa entre a geração térmica e geração hidráulica, para cada mês e para cada estado do sistema hidroelétrico, pode ser resolvida de forma a minimizar o valor esperado do custo de combustível e dos déficits que ocorrerão no horizonte de planejamento.

## 2- IDENTIFICAÇÃO DO MODELO

A disponibilidade de muitas séries sintéticas pode dar a impressão de que é possível ter confiança absoluta nos resultados encontrados com o seu emprego. É importante ressaltar, no entanto, que a geração de séries sintéticas não cria informação alguma. Existem até casos em que é preferível utilizar apenas a série histórica (Staschus e Kelman, 1987).

O processo de seleção de um modelo e de estimação de seus parâmetros é feito a partir de inferência com a série histórica. Portanto, a utilização do modelo estocástico de vazões deve ser encarada simplesmente como uma das maneiras de se manipular a informação existente. Nenhuma \* nova informação, é oriada, a não ser pela adoção de hipóteses subjacentes ao modelo estocástico. Estas hipóteses devem ser exaustivamente testadas antes que se aceite qualquer modelo. Infelizmente nunca se pode ter absoluta certeza quanto a identidade do verdadeiro modelo utilizado pela Natureza.

A primeira etapa na escolha do modelo estocástico, a identificação do modelo, é muito dependente da experiência do hidrólogo. Trata-se de escolher o modelo que seja mais indicado para um particular rio ou bacia hidrográfica. Para facilitar a discussão, imagine que apenas os dois modelos seguintes estejam sendo considerados para ajuste à série histórica de um rio hipotético.

$$\text{Modelo 1: } Y_{t+1} = \mu + \sigma Z_{t+1}$$

$$\text{Modelo 2: } Y_{t+1} = \mu + \rho(Y_t - \mu) + \sigma \sqrt{1-\rho^2} Z_{t+1}$$

onde

$Y_t$  = vazão no ano  $t$

$\mu$  =  $E(Y_t)$

$\sigma$  =  $\text{var}(Y_t)$

$\rho$  =  $\text{corr}(Y_t, Y_{t+1})$

$Z_{t+1}$  = "ruído" normal padrão. Isto é,  $Z_{t+1} \sim N(0,1)$ ,  $\forall t$  e

$Z_i$  independente de  $Z_j$  para  $i \neq j$ .

A etapa de identificação neste exemplo consiste em decidir qual das duas formulações é a mais indicada.

O Modelo 1 é um caso particular do Modelo 2, para  $p=0$ .

Como não existem razões físicas para um valor de  $p$  negativo, esta hipótese não é sequer cogitada. Por outro lado, valores positivos para  $p$  são perfeitamente aceitáveis, devido ao armazenamento de água no sub-solo. \*

Mesmo quando as afluências ao "reservatório" existente no sub-solo sejam independente em anos consecutivos, os estados de armazenamento poderão ser dependentes, desde que a capacidade do "reservatório" permita regularização pluri-anual. Sabe-se que o escoamento básico varia diretamente com o estado do armazenamento subterrâneo. Este é um conceito adotado por um grande número de modelos de transformação de chuva em vazão, que pode ser verificado empiricamente pela simples observação do grande número de nascentes que secam após uma prolongada estiagem.

Consequentemente o escoamento básico poderá ser dependente em anos consecutivos, mesmo que os processos meteorológicos não sejam dotados de suficiente "memória" para fazer com que a precipitação anual, e também o total infiltrado, sejam variáveis dependentes. Como o deflúvio anual pode ser conceitualmente separado em escoamento superficial e básico, pode-se dizer que a escolha do modelo matemático mais plausível, entre os dois candidatos acima especificados, depende da característica geológica (capacidade do armazenamento subterrâneo) da bacia hidrográfica sendo estudada.

Idealmente as informações sobre as características do solo, sub-solo e clima deveriam ser consideradas não apenas para a identificação do modelo estocástico mais apropriado, como também para a estimação dos parâmetros. A hidrologia, no entanto, não tem ainda uma metodologia capaz de atingir na prática este objetivo, apesar de inúmeros esforços feitos nesta direção (veja por exemplo Koch, 1985). Com base apenas na série histórica de vazões anuais, a discriminação entre os dois modelos pode ser feita através do seguinte teste de

hipótese:

Hipótese nula:  $\rho=0$

Hipótese alternativa:  $\rho>0$

Se a hipótese nula puder ser rejeitada, o Modelo 1 deve ser também rejeitado, possivelmente em favor do Modelo 2. O coeficiente de correlação amostral  $\hat{\rho}$  é a estatística indicada para a realização do teste.

Como se sabe, a não-rejeição da hipótese nula não equivale a sua aceitação. Isto é, é desconhecida a probabilidade de engano (erro tipo II) caso o Modelo 1 seja "aceito" com base na não-rejeição da hipótese nula. Por outro lado, caso a hipótese nula seja rejeitada, a probabilidade de se cometer um engano (erro tipo I) é igual ao nível de significância, previamente selecionado. Isto quer dizer que o teste só fornece um resultado afirmativo caso resulte na rejeição do Modelo 1. Caso contrário, ambos os modelos permanecem na "competição".

Um critério possível para discriminar entre dois ou mais modelos competidores é verificar qual a verossimilhança atribuída por cada um deles à série histórica: deve-se escolher o modelo para o qual a série histórica seja uma "típica" realização do processo estocástico. Convém lembrar que o método da máxima verossimilhança para estimação de parâmetros se baseia num princípio análogo. Este procedimento não pode ser empregado quando os modelos competidores tiverem número diferente de parâmetros porque o modelo com mais parâmetros tenderá a se ajustar melhor aos dados, e conseqüentemente produzir uma maior verossimilhança, sem que necessariamente seja melhor. Na realidade deve-se utilizar modelos com o menor número possível de parâmetros, para evitar que flutuações amostrais da série histórica sejam consideradas propriedades de população. Este é o chamado "princípio da parcimônia" (Box e Jenkins, 1976).

O critério de Akaike é uma medida que procura balancear a

meta de parcimônia de parâmetros com o objetivo de se obter um bom ajuste, medida pela função de verossimilhança:

$$A = -2 \ln (\text{máxima verossimilhança}) + 2 (\text{número de parâmetros})$$

Segundo o critério, deve ser selecionado o modelo com o menor valor de A. Recomenda a prudência que este critério não seja utilizado isoladamente. Afinal é possível que dois modelos com valores de A muito próximos resultem em decisões de Engenharia bastante distintas. Ao selecionar um modelo, o hidrólogo deve ter em mente as incertezas na estimação dos parâmetros (item 3) e, principalmente, os resultados dos testes de validação (item 5). Caso não haja uma clara definição em favor de um dos modelos competidores, a decisão final deverá seguir a tradição da Engenharia de adotar a solução que fique "em favor da segurança".

Box e Jenkins (1976) apresentaram diversos procedimentos mais elaborados para a identificação de modelos, aplicáveis à séries temporais pertencentes a família ARMA e Hipel e Maleod (1986) ampliaram a metodologia para aplicação específica à séries hidrológicas.

### 3- ESTIMAÇÃO DOS PARÂMETROS

Mesmo que o verdadeiro modelo fosse conhecido, restaria ainda estimar os parâmetros a partir da única informação disponível: a série histórica.

Na prática, grandezas amostrais, calculadas frequentemente a partir de registros curtos, são utilizadas em lugar dos valores de população. A geração de uma série sintética suficientemente longa faz com que as grandezas amostrais da série sintética convirjam para as grandezas amostrais da série histórica e não para as grandezas populacionais, como seria desejável. Isto é, os erros amostrais da série histórica são perpetuados na série

sintética. Este "defeito" do Metodo de Monte Carlos é chamado de tendenciosidade operacional.

Uma maneira de levar em consideração a incerteza quanto ao valor numérico dos parâmetros consiste em reconhecer que os estimadores dos parâmetros são também variáveis aleatórias..A geração de K séries sintéticas se processa segundo o algoritmo, onde  $k = a \times b$ :

```
Repita    i= 1,a
```

```
Sortele um vetor de parâmetros, digamos  $\theta_i$ , da
distribuição de probabilidades do estimador  $\hat{\theta}$ ,
digamos  $F_{\hat{\theta}}(\cdot)$  ;
```

```
Repita    j= 1,b
```

```
Sortele uma série sintética de vazões, dada o
parâmetro  $\theta_i$ 
```

Raramente a caracterização de  $F_{\hat{\theta}}(\cdot)$  é trivial. Quase sempre procedimentos mais sofisticados seriam necessários. Uma maneira de evitar esta complicação consiste em sortear  $\theta_i$  com a ajuda do modelo estocástico, de acordo com o algoritmo:

Ajuste o modelo estocástico à série histórica. Obtendo  $\theta_h$

Repita  $i=1, a$

Gere uma série sintética de comprimento igual ao da série histórica, dado  $\theta_h$ . Esta será uma série "filha"

Ajuste o modelo à série filha. Obtenha  $\theta_i$

Repita  $j=1, b$

Gere uma série sintética de comprimento não necessariamente igual ao da série histórica. Esta será uma série "neta"

Com esta abordagem, a incerteza quanto aos parâmetros de população é traduzido pela variação nos parâmetros das séries filhas, uma vez que cada série filha tem o mesmo comprimento da série histórica.

#### 4- VERIFICAÇÃO DO MODELO

É necessário confirmar que as hipóteses subjacentes ao modelo estocástico "candidato à adoção" são satisfeitas, antes que efetivamente se utilize o modelo. Por exemplo, suponha que o Modelo 2 seja o candidato. Existem duas hipóteses associadas a este modelo:

- a)  $Z_t$  tem distribuição normal padrão. Para testar esta hipótese é preciso, antes de mais nada, obter um conjunto de  $n-1$  "observações"  $(z_2, z_3, \dots, z_n)$  através da solução da equação do modelo para  $Z$ :

$$Z_{t+1} = [Y_{t+1} - \mu - \rho(Y_t - \mu)] / (\sigma \sqrt{1 - \rho^2})$$

O teste de hipótese propriamente dito pode ser feito de diversas maneiras.

b)  $Z_i$  independente de  $Z_j$ , para  $i \neq j$ . Para testar esta hipótese é necessário que se calcule os coeficientes de correlação amostrais  $c\hat{o}rr(Z_t, Z_{t+k})$ . Se a hipótese de independência estiver correta, o correlograma tenderá a ser uma linha horizontal, perto de zero. \*

O teste de hipótese propriamente dito pode ser feito de diversas maneiras, como por exemplo através do teste de portmanteau.

Se qualquer das hipótese subjacentes ao modelo puder ser rejeitada, o modelo deverá também ser rejeitado.

#### 5- VALIDAÇÃO DO MODELO

A confiança que se possa ter num modelo estocástico de vazões depende da capacidade que ele tenha de preservar nas séries sintéticas algumas propriedades observadas na série histórica. Uma "propriedade" pode ser constituída por um único valor, como por exemplo a vazão média observada em  $n$  anos, ou pode ser um conjunto de valores, como por exemplo a distribuição empírica de probabilidades da máxima vazão observada em cada ano. Diz-se que um modelo preserva uma determinada propriedade quando não se pode distinguir estatisticamente a série histórica da sintética, com base nas observações desta propriedade nas duas séries.

Quando alguma propriedade é utilizada para a determinação de um parâmetro do modelo, esta propriedade é automaticamente preservada, "por construção". Neste caso comparar a observação da propriedade da série sintética com seu correspondente da série histórica serve apenas para verificar a adequação do programa de computador utilizado, e não para validar o modelo (Kelman e Pereira, 1977, Stedinger e Taylor, 1982).

A seleção do conjunto de propriedade em que se vai focar

atenção depende da aplicação que se pretenda dar às séries sintéticas. Por exemplo, quando o planejamento da expansão do sistema hidrelétrico é feito com auxílio de séries sintéticas mensais, a atenção deve estar voltada primordialmente para a preservação dos volumes afluentes durante as secas (Kelman e Pereira, 1977). Quando o cálculo do volume de amortecimento de cheias é feito com auxílio de séries sintéticas diárias, a atenção deve estar voltada para a preservação dos picos e volumes das cheias excepcionais.

Suponha que a partir da série histórica e da série sintética produzida por algum modelo estocástico seja possível extrair amostras aleatórias de alguma variável cujas propriedades haja interesse em preservar, respectivamente  $\{x_h(1), x_h(2), \dots, x_h(j)\}$  e  $\{x_s(1), x_s(2), \dots, x_s(k)\}$ . Como a série sintética é em geral mais longa do que a histórica, é de se supor que  $k > j$ . Por exemplo se  $X$  for a máxima vazão diária num período de um ano, e admitindo que o modelo de vazões diárias sob escrutínio não use esta variável como parâmetro,  $j$  e  $k$  serão respectivamente o número de anos na série histórica e sintética. A validação do modelo é feita testando a hipótese de que ambas as amostras tenham sido sorteadas de uma mesma população. Se a hipótese nula for rejeitada, o modelo deverá também ser rejeitado. Como em princípio não se sabe qual a forma da distribuição de probabilidade de  $X$ , os métodos não-paramétricos são particularmente indicados (Bradley, 1968). Por exemplo, o teste de Smirnov pede pelo cálculo da maior diferença entre as distribuições empíricas provenientes das duas amostras.

Suponha que a partir da série histórica seja possível extrair apenas uma observação da variável cujas propriedades haja interesse em preservar  $x_h$ . Por exemplo,  $X$  pode ser o tamanho do reservatório capaz de garantir um determinado deflúvio anual durante  $n$  anos (duração da série histórica). O modelo estocástico "candidato" pode ser utilizado para produzir  $r$  séries sintéticas, cada uma delas também

de comprimento  $n$ , resultando na amostra  $\{x_s(1), x_s(2), \dots, x_s(r)\}$ . Uma proporção muito grande ou muito pequena de elementos do conjunto que sejam menores do que  $x_h$  indica que a série histórica é "atípica" para o modelo considerado, o que sugere a rejeição do modelo.\* Mais formalmente o que se está estimando é a probabilidade  $\beta = P(X_s < x_h)$ .

O leitor notará um paradoxo neste procedimento: um dos objetivos da Hidrologia Estocástica é o de estimar a probabilidade  $\beta$ . No entanto, quando a estimativa for menor do que 0,05 ou maior do que 0,95, que são valores legítimos, o teste de validação indicará a rejeição do modelo. A analogia com o conceito de erro tipo I em testes de hipóteses é evidente: em 10% das vezes em que o modelo for adequado, o teste de validação recomendará rejeição. É exatamente nestes 10% de casos que o uso do modelo estocástico tem mais chances de causar importantes consequências no processo decisório. Por isto, é conveniente que se avalie um modelo estocástico em relação a diversas variáveis aleatórias e não apenas especificamente em relação a variável que fornece uma resposta a um problema de engenharia. É com base no conjunto dos resultados dos testes de validação que se deve tomar uma decisão quanto a rejeição ou não do modelo estocástico.

Uma variante do procedimento de validação foi sugerida por Thompstone et al (1987), que recomendaram a adoção de amostras particionadas. Nesta abordagem ajusta-se o modelo estocástico a apenas os primeiros  $n_1$  anos da série histórica, reservando-se os  $n_2$  anos remanescentes ( $n_1 + n_2 = n$ ) para o cálculo de  $x_h$ . O teste é feito com  $r$  séries sintéticas de comprimento igual a  $n_2$  anos. Esta variante resulta em resultados mais confiáveis do que se conseguiria ajustando o modelo estocástico à totalidade da série histórica.

## REFERÊNCIAS

- (1) Box, G.E.P. e G.Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, 1976. \*
- (2) Bradley, J.V., Distribution - Free Statistical Tests, Prentice-Hall, Inc, 1968.
- (3) Hipel, K.W. e A.I. Mc.Leod, Time Series Modelling for Water Resources and Environmental Engineers, Elsevier, 1986.
- (4) Kelman J., M.V.F.Pereira, Critérios de Avaliação para Modelos de Séries Hidrológicas, IV Seminário Nacional de Produção e Transmissão de Energia Elétrica, Rio de Janeiro, 1977.
- (5) Koch, R.W., A Stochastic Streamflow Model Based on Physical Principles, Water Resources Research, 21(4), 1985.
- (6) Staschus, K., J.Kelman, A Statistical Approach to Determine the Benefits of Synthetic Hydrology: A Case Study, artigo em elaboração, 1987.
- (7) Stedinger, J.R., M.R.Taylor, Synthetic Streamflow Generation 1. Model Verification and Validation, Water Resources Research, 18(4), 1982.

(8) Thompstone, R.M. K.W. Hipel, A.I. Mc.Leod, Monthly Hydrological Time Series, in Advances in the Statistical Sciences No 37, Stochastic Hydrology, D.Reidel Publishing Co 1987.

\*

(9) Terry, L.A., M.V.F.Pereira, T.A. Araripe Neto, L.F.C. Silva, P.R.H. Sales, Coordination the Energy Generation of the Brazilian National Hydrothermal Electrical Generating System, Interfaces - 16:1. January-February 1986.

f