

MODELAGEM DE SÉRIES HIDROLÓGICAS ATRAVÉS DE MODELOS DE MEMÓRIA LONGA

Guilherme Armando de Almeida Pereira

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil - 22451-900
gaap@ele.puc-rio.br

Reinaldo Castro Souza 88151

Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica
Rua Marquês de São Vicente, 225, Gávea - Rio de Janeiro, RJ - Brasil - 22451-900
reinaldo@ele.puc-rio.br

RESUMO

O planejamento da operação do sistema hidrotérmico brasileiro é realizado através de modelos de otimização que utilizam como parâmetros séries hidrológicas sintéticas geradas a partir do modelo estatístico *PAR (p)* (periódico autorregressivo). Recentemente, iniciaram-se as buscas por modelos alternativos capazes de captar efeitos que o tradicional modelo *PAR (p)* não incorpora. Dentre estes, destacam-se os fenômenos cíclicos e/ou os efeitos de memória longa. Memória longa em uma série temporal pode ser definida como a dependência significativa entre *lags* separados por um distante período de tempo. Desta maneira, este trabalho tem como objetivo ajustar modelos de memória longa capazes de gerar séries hidrológicas sintéticas a partir de séries históricas de energia natural afluenta. Através das métricas empregadas, pode-se concluir que a metodologia empregada é capaz de gerar boas séries sintéticas.

PALAVARAS CHAVE. Séries temporais com memória longa, modelos estocásticos, cenários hidrológicos.

EN – PO na Área de Energia; MP – Modelos Probabilísticos

ABSTRACT

The operation planning of the Brazilian hydrothermal system is accomplished through optimization models that use inputs such as synthetic hydrologic series generated from the statistical model *PAR (p)* (autoregressive periodic). Recently, there has been began the search for alternative models able to capture the effects that the traditional model *PAR (p)* does not incorporate. Chief among these are natural phenomena and/or long memory effects. Long memory in a time series can be defined as a significant dependence between lags separated by a far period of time. Thus, this paper aims to set long memory models capable of generating synthetic hydrologic series from time series of natural energy affluent. Through the metric used, it can be concluded that the method is able to generate valuable synthetic series.

KEYWORDS. Time series with long memory, stochastic models, hydrological scenarios.

EN – OP in Energy Area; PM – Probabilistic Models

1 – Introdução

Sabe-se que, no Brasil, mesmo com a crescente diversificação da matriz energética, aproximadamente 85% do potencial de geração da energia é oriunda de usinas hidrelétricas. Uma das principais características de matrizes energéticas com esta composição é a forte dependência dos regimes de chuvas. Este fato gera uma incerteza na operacionalização de tal matriz tornando assim o seu planejamento não trivial.

Cabe ao Operador Nacional do Sistema Elétrico (ONS) a responsabilidade do planejamento e operacionalização do sistema elétrico brasileiro. Nessa atividade, segundo (Fayal, 2008), são adotados modelos de simulação e otimização que utilizam a vazão natural prevista e/ou verificada como entradas para obter resultados que indiquem a mais adequada situação de armazenamento, de defluência e de geração dos aproveitamentos hidroelétricos, a cada instante de tempo.

Desta forma, estimar modelos que sejam capazes de prever e/ou simular as séries hidrológicas existentes também é de extrema importância para o planejamento ótimo do sistema. É de conhecimento comum que pequenos avanços em tais modelos são capazes de possibilitar a melhoria do planejamento da operação do sistema, a qual é diretamente revertida em economia de investimentos, modicidade tarifária e melhor aproveitamento dos recursos disponíveis no sistema. Isso justifica o alto investimento que vem sendo realizado pelo setor.

Para o planejamento da operação de médio prazo no Brasil, a ferramenta computacional utilizada é o modelo NEWAVE, que segundo (Oliveira, 2010), o planejamento é representado por um problema de programação linear estocástica multi-estágio cujo objetivo é a minimização do custo total de operação.

Para que isso seja realizado, um dos principais parâmetros do modelo é um conjunto de séries hidrológicas sintéticas geradas a partir do histórico de cada um dos 4 subsistemas brasileiros (Sudeste/Centro-Oeste, Sul, Nordeste e Norte). Para tal geração, o modelo estatístico utilizado é uma extensão dos modelos $ARMA(p, q)$ denominado $PAR(p)$ (autorregressivo periódico), (Thomas & Fiering, 1962). O $PAR(p)$ é utilizado em séries históricas que apresentam uma estrutura de autocorrelação que não depende somente do intervalo de tempo entre as observações, mas também do período observado. Desta forma, para cada período um modelo $AR(p)$ é ajustado, ou seja, se a série é mensal, 12 modelos $AR(p)$ são ajustados cada uma com uma ordem p não necessariamente igual. O $PAR(p)$ vem sendo amplamente utilizado para a geração de séries sintéticas, entretanto, recentemente, iniciou-se a busca por novas ferramentas estatísticas capazes de gerar séries sintéticas de energia afluentes.

Utilizar métodos capazes de captar os efeitos que o modelo $PAR(p)$ não consegue obter é uma das razões da busca por novos modelos geradores de cenários. Vale destacar, dentre estes efeitos, principalmente os de memória longa e/ou cíclicos.

Memória longa, ou persistência em uma série temporal, pode ser definida como a presença de dependência entre observações muito distantes no tempo, diferente de modelos tradicionais onde a correlação entre observações separadas por um longo período de tempo são consideradas nulas ou desprezíveis.

Vários são os modelos de memória longa existentes na literatura. O modelo utilizado neste trabalho é denominado $SARFIMA(p, d, q).(P, D, Q)_s$, (PORTER-HUDAK, 1990). Este modelo é uma extensão dos modelos $ARIMA$ de Box & Jenkins onde os parâmetros de diferenciação d e D podem assumir valores fracionários. Esta adaptação é realizada para que seja possível capturar efeitos de longo prazo e/ou ciclos existentes na série em estudo.

Portanto, este trabalho tem como principal a geração de séries hidrológicas sintéticas por meio de modelos de memória longa aplicados a séries de ENA (energia natural afluente) do subsistema Sul e do Subsistema Sudeste. As ENAs são computadas a partir das vazões naturais afluentes e das produtibilidades equivalentes ao armazenamento de 65% do volume útil dos reservatórios de aproveitamentos hidroelétricos, (Marcato, 2002).

Isto posto, este trabalho está dividido da seguinte maneira: na seção 2, empreende-se-á uma revisão de literatura dos modelos *SARFIMA*. Na seção 3, será exposto o *Bootstrap* e como ele foi empregado na construção de intervalos de confiança e na geração de séries sintéticas. Ao longo da quarta seção, serão apresentados os critérios de avaliação utilizados para a validação dos cenários hidrológicos sintéticos. A seção 5 consiste na apresentação do modelo ajustado, assim como os cenários gerados e avaliados. No último capítulo, serão feitas as considerações finais, bem como propostas de novos trabalhos.

2. Metodologia

2.1 – O Modelo SARFIMA

O fenômeno de memória longa foi primeiramente observado em séries hidrológicas e climatológicas no início dos anos de 1950, embora apenas na década 80 é que econometristas e estatísticos começaram a dar a devida atenção a este tipo de fenômeno.

Persistência ou longa dependência é o termo usado para descrever séries temporais cuja função de autocorrelação exibe um comportamento que não é compatível nem com séries estacionárias, ou com séries não estacionárias.

Neste tipo de processo, as autocorrelações amostrais entre lags distantes não podem ser consideradas desprezíveis, diferentemente de processos de memória curta, processos *ARMA*, por exemplo, cujas observações separadas por um longo período de tempo são consideradas independentes.

No domínio do tempo, a função de autocorrelação não é absolutamente somável apresentando o seguinte comportamento:

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty \quad (1)$$

Outra característica é que as autocorrelações $\rho(k)$ tem um decaimento lento e hiperbolico, equação 2, diferentemente do decaimento rápido e exponencial, equação 3, observado em séries tradicionais.

$$\rho_k \sim a^k \quad (2)$$

onde $0 < a < 1$.

$$\rho(k) \sim e^{-|k|^{2d-1}}, |k| \rightarrow \infty \quad (3)$$

No domínio da frequência, a presença de longa dependência pode ser observada através da função espectral que se torna ilimitada conforme a frequência tende a zero:

$$f_X(\omega) \rightarrow \infty \quad \text{quando} \quad \omega \rightarrow 0 \quad (4)$$

Souza (1997) cita outras características deste tipo de série, dentre elas, a) períodos longos em que as observações tendem a se manter em um nível baixo ou elevado; b) ao se observar a série de perto, isto é, ao se observar poucos anos, é possível perceber aparentes ciclos e/ou tendências, entretanto, ao se olhar a série como um todo (série composta de todas as observações possíveis), nenhum ciclo ou tendência parece persistir; c) a série aparenta estacionariedade; d) o gráfico log do periodograma versus log das frequências fica disperso em torno de uma linha reta decrescente.

Para se conseguir modelar estas características (Granger & Joyeux, 1980) e (Hosking, 1981) desenvolveram o modelo *ARFIMA* (p, d, q). Este é uma generalização dos modelos *ARIMA* para o caso onde d assume qualquer valor real, sendo um dos modelos de memória longa mais flexível e abrangente.

Esta metodologia foi ampliada para séries temporais com sazonalidade. Uma série temporal exibe um fenômeno sazonal quando este se repete regularmente depois de período de tempo não superior a um ano. O menor período de tempo desta repetição do fenômeno é denominado período sazonal.

O estudo da componente sazonal considerando o fenômeno de longa dependência é bem recente.

(Peires & Singh, 1996) definem o processo *SARFIMA* $\{X_t\}$ como:

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)a_t \quad (5)$$

Os polinômios autorregressivos simples e sazonais, médias móveis simples e sazonais, são respectivamente, $\phi(B), \Phi(B^s), \theta(B), \Theta(B^s)$. As ordens dos polinômios p, q, P, Q assumem valores inteiros e positivos enquanto que os parâmetros fracionários d e D assumem valores reais.

O operador de diferenciação fracionária simples, $(1-B)^d$ e operador de diferenciação fracionária sazonal, $(1-B^s)^D$, pode ser visualizado em (6) e (7):

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d}{2!}(1-d)B^2 - \dots \quad (6)$$

$$(1-B^s)^D = 1 - DB^s - \frac{D(1-D)}{2!}B^{2s} - \frac{D(1-D)(2-D)}{3!}B^{3s} - \dots \quad (7)$$

O modelo é estacionário e invertível se e somente se, $|d+D| < 1/2$, $|D| < 1/2$, $|d| < 1/2$ e $\phi(B), \theta(B), \Phi(B)$ e $\Theta(B)$ têm suas raízes fora do círculo unitário. O processo possui características de memória longa quando $|d+D| > 0$. A função densidade espectral do modelo da equação 8 é dada por:

$$f_x(\omega) = f_u(\omega) [2\text{sen}(\omega s / 2)]^{-2D} [2\text{sen}(\omega / 2)]^{-2d} \quad (8)$$

para $-\pi \leq \omega \leq \pi$, em que $f_u(\cdot)$ é a densidade espectral do processo *SARMA*.

Como dito inicialmente, no modelo *SARFIMA*, d e D são valores fracionários que incorporam os efeitos de memória longa ao modelo. Para que isto seja possível, é necessária a estimação dos mesmos, uma vez que estes parâmetros são desconhecidos.

O método de estimação utilizado foi similar ao método sugerido por (Reisen, Rodrigues & Palma, 2006a, b).

Este método de estimação consiste em uma regressão usando a função densidade espectral (equação 8) para montar uma regressão entre o periodograma e as frequências harmônicas.

Para isto, deve-se logaritmizar a equação 8, obtendo assim:

$$\ln f_x(\omega) = \ln f_u(\omega) - D \ln [2\text{sen}(\omega s / 2)]^2 - d \ln [2\text{sen}(\omega / 2)]^2 \quad (9)$$

A equação 9 é similar a uma regressão linear múltipla da forma:

$$Y_j = a + b_1 X_{1j} + b_2 X_{2j} + \varepsilon_j \quad (10)$$

em que, $Y_j = \ln I(\omega_j)$, $X_{1j} = \ln [2\text{sen}(\omega_j s)]^2$, $X_{2j} = \ln [2\text{sen}(\omega_j / 2)]^2$, $b_1 = -D$, $b_2 = -d$ e

$$\varepsilon_j = \ln \frac{I(\omega_j)}{f(\omega_j)} - E \left[\ln \frac{I(\omega_j)}{f(\omega_j)} \right].$$

Desta maneira, $\hat{d}$ e $\hat{D}$ devem estimados através de mínimos quadrados ordinários. Como pode ser observada, a estimativa dos parâmetros fracionários é o valor encontrado para os coeficientes da regressão (b_1 e b_2) multiplicado por menos um.

Cabe salientar também que o número de observações na regressão é determinado por $g(n) = n^\alpha$, em que α é uma constante entre zero e um e n é o tamanho da série temporal. Além disto, $g(n)$ deve satisfazer a condição $\left(\frac{g(n)}{n} \right) \log g(n) + \frac{1}{g(n)} \rightarrow 0$ quando $g(n) \rightarrow \infty$ e $n \rightarrow \infty$. Em termos práticos, pode-se dizer que a regressão é realizada com $j = 1, 2, \dots, K, g(n)$.

Como estimativa da função densidade espectral, $I(\omega)$, neste trabalho foi utilizado a função periodograma suavizada expressa em (11).

$$f_{sp}(\omega) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{n-1} \lambda(s) \gamma(s) \cos(s\omega) \quad (11)$$

Um dos motivos da escolha da função periodograma suavizado, onde $\lambda(s)$ é a janela espectral ou função de suavização, como estimador da função espectral, é que o periodograma não é uma estimativa consistente do espectro. A janela empregada será a de Parzen:

$$\lambda(s) = \begin{cases} 1 - 6(s/m)^2 + 6(s/m)^3, & |s| < m/2, \\ 2(1 - (s/m))^3, & m/2 \leq |s| \leq m, \\ 0, & |s| > m \end{cases} \quad (12)$$

onde $m = n^\beta$ e $0 < \beta < 1$. Várias são as janelas existentes na literatura, dentre elas pode-se citar a janela de Daniell, janela de Tukey e janela de Bartlett. A razão de se escolher a janela de Parzen é que esta tem a característica de não produzir estimativas negativas da função densidade espectral.

2.2 – Bootstrap em modelos de Memória longa para geração de séries sintéticas

O Bootstrap é uma técnica estatística não paramétrica computacionalmente intensiva de reamostragem, introduzido por (Efron, 1979), e tem como finalidade obter informações de características da distribuição de alguma variável aleatória. Para isto, aproxima-se uma distribuição de probabilidade através de uma função empírica obtida de uma amostra finita. Normalmente, esta técnica é empregada quando a distribuição de interesse é de difícil, ou até impossível avaliação analítica ou quando só a teoria assintótica está disponível.

Segundo (Souza & Camargo, 2004), o *bootstrap* pode ser realizado de duas maneiras em séries temporais: o *bootstrap* nos resíduos do modelo ou o *moving blocks*. Para o primeiro caso, é necessário o ajuste de um modelo probabilístico para que os resíduos obtidos, que são independentes, possam ser utilizados na obtenção de outras séries *bootstrapadas*. Este é o método mais difundido na literatura. Este método é conhecido como *bootstrap* condicional, uma vez que a utilização dele está condicionado a um modelo previamente ajustado. O segundo método consiste em construir blocos de tamanho “M” da série original e posteriormente realizar o sorteio com reposição desses blocos, até se formar uma amostra *bootstrap*. Esta abordagem é denominada *bootstrap* incondicional, uma vez que para a utilização não é necessário a estimação de nenhum modelo.

O *bootstrap* será empregado neste trabalho em duas fases distintas: geração de séries hidrológicas sintéticas e seleção de cenários.

Um modelo que descreva a estrutura de probabilidade de uma sequência de observações é chamado de processo estocástico. Processos estocásticos são sistemas que evoluem no tempo e/ou no espaço, de acordo com leis probabilísticas.

Uma série temporal, ou série histórica, nada mais é do que apenas uma das possíveis realizações de um processo estocástico. Ao se gerar séries sintéticas através de um modelo ajustado à série histórica, estar-se-á tentando reproduzir novas realizações desse processo. Isto é, estar-se-á gerando quantas séries se desejar igualmente prováveis do ponto de vista estatístico, porém diferente do histórico.

A primeira aplicação do *bootstrap* neste trabalho está relacionada à geração de séries sintéticas com base no histórico de energia natural afluyente disponível (ENA). Para a geração de cenários, o *bootstrap* será realizado nos resíduos do modelo.

Em linhas gerais, com base em um modelo ajustado, realiza-se sorteios aleatórios com reposição dos resíduos e para cada erro sorteado, um novo ponto da série é gerado.

A equação do modelo pode ser obtida resolvendo a equação de diferenças, expressa em 5. Como pode ser observado na equação supracitada, existem dois polinômios de ordem infinita.

Em termos práticos, quando se tem uma série histórica com n observações, utilizam-se somente os K primeiros termos desse polinômio, com $K \leq n$.

Como descrito anteriormente, e adotando K igual a 936, uma vez que será utilizado todo o histórico disponível para geração, a equação do modelo é dada por:

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + K + \alpha_{935} Z_{t-935} + \varepsilon_{t,m}^* \quad (15)$$

Note que o termo ε possui o subscrito m . O Subscrito m denota o mês referente ao ponto que estará sendo gerado. Dessa forma, para se gerar um ponto referente a janeiro, um erro corresponde àquele mês deverá ser sorteado. De outra forma, os 936 erros do modelo foram separados em 12 vetores de tamanho 78. Assim sendo, ao se gerar um ponto de janeiro, sorteia-se um erro do vetor de erros de janeiro.

Ao se utilizar o *bootstrap* para gerar séries hidrológicas sintéticas, embora todas as séries geradas sejam igualmente possíveis do ponto de vista estatístico, nem todas as séries são plausíveis do ponto de vista físico. Neste trabalho, séries com valores negativos não podem ser considerada dada a característica da variável em questão.

Uma das alternativas possíveis ao problema de geração de valores negativos é gerar um número de cenários maior do que o necessário e excluir cenários com valores negativos. Todavia, ao se proceder desta maneira problemas foram encontrados.

Ao se realizar os testes para verificar a adequabilidade entre as séries geradas (com vazões negativas e positivas) e o histórico, os resultados dos testes indicavam que as séries sintéticas eram estatisticamente similares ao histórico. Entretanto, ao se retirar as séries com vazões negativas e realizar os mesmos testes novamente, foi verificado uma elevação da média dos cenários gerados em relação às estatísticas do histórico. Isto aconteceu porque existiam cenários com aflúências baixas, cenários com aflúências similares ao histórico e cenários com aflúências altas. Ao se realizar os testes estatísticos descritos, os resultados eram bons, principalmente porque as médias eram estatisticamente iguais. Retirando os cenários com aflúências negativas, em sua maioria correspondente aos cenários com aflúências baixas, a média desta nova matriz era superior ao histórico (principalmente porque existiam cenários com médias muito elevadas) e, portanto, os testes empregados apresentaram resultados ruins. Este fato está ilustrado na figura 1 e figura 2.

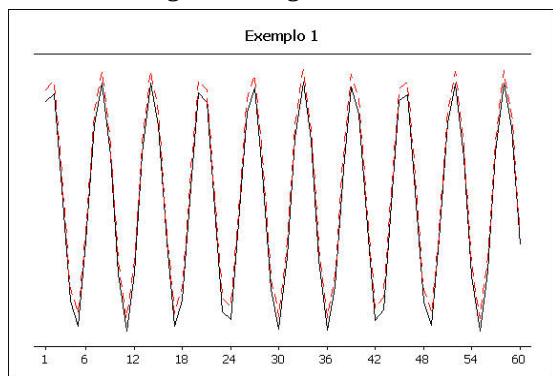


Figura 1 – Média histórica e média dos cenários (valores positivos e negativos)

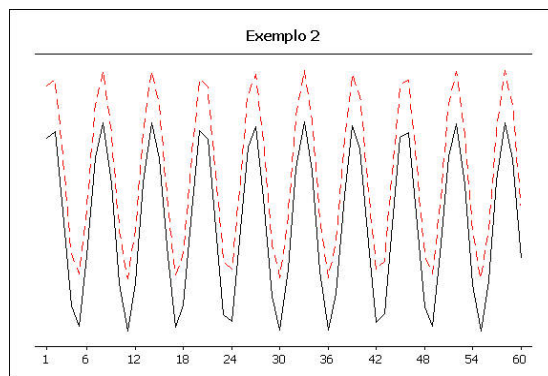


Figura 2 – Média histórica e média dos cenários (somente valores positivos)

Para solucionar este problema, uma adaptação foi realizada no método para que se fosse possível obter séries sintéticas com médias fidedignas ao histórico.

Esta nova abordagem consiste em estipular limites para a média. O objetivo é excluir cenários com médias muito diferentes do histórico disponível, cenários estes, que deturpam significativamente a análise.

Para a determinação destes limites, foi utilizado o *bootstrap* com *moving blocks* no histórico de ENA. O procedimento de *moving blocks* foi inicialmente proposto por (Efron & Tibshirani, 1993) e consiste em dividir blocos de tamanho " M " da série original. Estes blocos

são reamostrados com reposição até que se construa uma série do mesmo tamanho da original. Este processo é realizado B vezes gerando, assim, B amostras *bootstraps*.

De maneira resumida, pode-se dizer que o *bootstrap* será realizado neste trabalho em duas fases distintas como pode ser observado na tabela 1 abaixo:

Tabela 1 – Aplicações do *Bootstrap*

Tipo	Aplicação
<i>Bootstrap</i> nos resíduos do modelo de séries temporais	Geração de cenários hidrológicos sintéticos
<i>Bootstrap</i> Moving Blocks	Intervalos para seleção de cenários

2.3 Avaliações dos Cenários

Como dito anteriormente, o objetivo deste trabalho é a geração de cenários hidrológicos sintéticos a partir de diferentes modelos. É desejável que um modelo de geração de séries sintéticas preserve as principais características da série histórica. Isto quer dizer que a utilidade de um modelo pode ser verificada pela sua habilidade de reproduzir distribuições de probabilidade de variáveis aleatórias relevantes ao processo.

Desta maneira, serão apresentados os testes para verificar a qualidade dos cenários sintéticos gerados. Foram realizados o teste de média, o teste de variância, o teste de kolmogorov-Sminorv e a análise de sequência.

O teste de média tem como objetivo verificar se os cenários gerados possuem a mesma média do histórico disponível. A análise é realizada periodicamente, ou seja, será verificado se cada período gerado (janeiro, fevereiro,..., dezembro) é estatisticamente igual ao respectivo período no histórico.

A hipótese nula e alternativa são:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_1 : \mu_1 &\neq \mu_2 \end{aligned} \quad (6)$$

O teste para verificar se a variância dos cenários gerados são estatisticamente iguais ao a variância do histórico foi o teste de Levene onde as hipóteses nulas e alternativas são respectivamente:

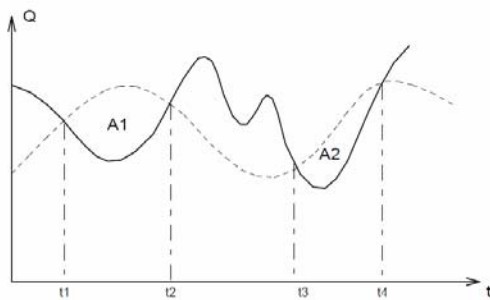
$$\begin{aligned} H_0 : \sigma_1^2 &= \dots = \sigma_k^2 \\ H_1 : \sigma_i^2 &\neq \sigma_j^2 \end{aligned} \quad (7)$$

para $i = 1, \dots, k$ e $j = 1, \dots, k$ e para algum $i \neq j$

O terceiro teste realizado foi o teste de Kolmogorov-Sminorv e tem como objetivo verificar se a distribuição de probabilidade dos cenários sintéticos é igual à distribuição da série original. A hipótese nula e alternativa são:

$$\begin{aligned} H_0 : F_1(x) &= F_2(x) \\ H_1 : F_1(x) &\neq F_2(x) \end{aligned} \quad (8)$$

Complementando a validação dos cenários, foi realizada a análise de sequência. Para isto, é necessário o conceito de sequência negativa. Uma sequência negativa é definida por um período de tempo em que as vazões estão continuamente abaixo de valores preestabelecidos (média do histórico). Com as sequências calculadas, é possível criar três variáveis (*comprimento*, *soma* e *intensidade de sequência*). A figura 5 abaixo ilustra o conceito de sequência negativa e como as variáveis são criadas a partir dela.



Variável Aleatória	Cálculo
Comprimento	$C = (t_2 - t_1)$
Soma	$S = \sum_{i=t_1}^{t_2} (Z_i - \mu_i)$
Intensidade	$I = \frac{S}{C} = \frac{\sum_{i=t_1}^{t_2} (Z_i - \mu_i)}{(t_2 - t_1)}$

Figura 3 – Análise de Sequências Negativas

Realizando os cálculos para cada sequência negativa encontrada, tanto para os cenários gerados, quanto para o histórico disponível, é possível obter assim amostras de cada variável. Assim sendo, tanto para o histórico, quanto para os cenários sintéticos foram obtidas três variáveis aleatórias.

Isto posto, em posse de duas amostras de cada variável é possível testar hipóteses se as amostras são provenientes da mesma distribuição por meio dos testes estatísticos de aderência. A variável *comprimento de sequência* é avaliada pelo teste Qui-Quadrado, enquanto que as variáveis *soma* e *intensidade de sequência* serão avaliadas pelo teste de Kolmogorov-Sminorv. Analogamente, também foi realizado a análise de sequência positiva. Uma sequência positiva é definida por um período de tempo em que as vazões estão continuamente acima de valores preestabelecidos.

3. Resultados

Nesta seção serão apresentados os resultados obtidos. As séries utilizadas são mensais com início em janeiro de 1931 e término em dezembro de 2008 totalizando 936 observações e os subsistemas analisados foram o subsistema Sudeste/Centro-Oeste e o subsistema Sul, pois dentre os quatro subsistemas estes foram onde a presença de memória longa é evidente, (Pereira, 2011).

A Energia Natural Afluente é computada a partir das vazões naturais e das produtibilidades equivalentes ao armazenamento de 65% do volume útil dos reservatórios de aproveitamentos hidroelétricos, (Marcato, 2002).

Na tabela 2, encontram-se os valores estimados para os parâmetros de memória longa d e D . Como pode ser observado, os parâmetros estimados satisfazem as condições de estacionariedade, invertibilidade e memória longa, tanto simples como sazonal. Para a definição de α e β onde $g(n) = n^\alpha$ (para regressão) e $m = n^\beta$ (janela de Parzen), foram testadas diversas combinações, variando ambos os parâmetros entre 0.6 e 0.9, utilizando tanto a função periodograma, quanto a função periodograma suavizado como estimador do espectro no método de regressão. Os valores destes hiperparâmetros foram escolhidos levando em consideração os resultados, na ótica da geração de cenários. Neste sentido, adotou-se a função periodograma suavizado com os valores de α e β iguais a 0,7.

Tabela 2 – Parâmetros fracionários do modelo SARFIMA

	d	D
Sudeste	0.1491	0.0736
Sul	0.1857	0.0203

Após a estimação dos parâmetros d e D , de acordo com método semiparamétrico de construção de modelos *SARFIMA*, deve-se diferenciar a série e estimar os parâmetros autorregressivos e médias móveis.

Os modelos selecionados foram o *SARFIMA* $(2, \hat{d}, 0).(1, \hat{D}, 0)_{12}$ para o subsistema Sudeste/Centro-Oeste e um *SARFIMA* $(1, \hat{d}, 0).(1, \hat{D}, 0)_{12}$ para o subsistema Sul.

3.1 – Seleção de Cenários

É necessário que em modelos de geração de séries hidrológicas sintéticas, os valores gerados sejam positivos e, ao se utilizar o *bootstrap*, esta exigência nem sempre é atendida.

Assim sendo, para que fosse possível obter séries apenas com afluências positivas, foi gerada uma matriz de cenários com tamanho superior ao necessário e desta matriz foi construída uma nova, selecionando apenas os cenários com valores positivos. Em seguida, retirou-se dessa matriz apenas os cenários com médias dentro dos limites preestabelecidos. Estes cenários selecionados é que fazem parte da matriz final que foi avaliada.

Para a construção desses intervalos, utilizou-se o *bootstrap* na série histórica. Foram geradas dez mil novas séries e para cada série foi calculada a média. Essas médias constituem a distribuição *bootstrap* da média do histórico. Os intervalos inferiores e superiores estipulados são iguais aos valores mínimos e máximos desta distribuição.

Nesta abordagem foi utilizado o *moving blocks* com blocos de tamanho três, para que a estrutura de autocorrelação e a estrutura hidrológica (períodos de secas e chuvas) fossem mantidas. Na tabela 3, são apresentados os limites inferiores e superiores para seleção de cenários.

Tabela 3 – Mínimo e Máximo para Seleção de Cenários

	Mínimo	Máximo
Sudeste	30596 MWmed	36766 MWmed
Sul	6914 MWmed	9509 MWmed

3.2 Geração de Cenários

Na figura 4 e 5, podem ser observados todos os cenários gerados, assim como a sua média mensal e a média mensal do histórico reproduzida para cinco anos. No total, foram selecionados 200 cenários de 5 anos, o que corresponde a 60 períodos gerados.

Através da análise gráfica, pode-se observar que as médias dos cenários sintéticos (linha preta pontilhada) são similares à média histórica (linha vermelha) que praticamente se sobrepõem. Também pode ser visto que a envoltória dos cenários sintéticos reflete corretamente os períodos hidrológicos, ou seja, os cenários reproduzem afluências altas em períodos chuvosos e baixas afluências em períodos secos.

Com relação aos testes estatísticos realizados, tabela 4, Toda a análise realizada levou em consideração um nível de significância de 5%.

Vale destacar que a análise realizada foi periódica entre os cenários gerados e o histórico, ou seja, procurou-se verificar se cada período gerado possui a variável de interesse estatisticamente igual ao período equivalente no histórico, isto é, se os janelos sintéticos são iguais aos janelos realizados e assim sucessivamente para todos os 60 períodos gerados.

Com relação ao teste de média, que é o mais importante, o índice de aprovação foi de 99% no subsistema Sul e 97% no subsistema Sudeste/Centro-Oeste. Isto mostra que o modelo proposto é capaz de reproduzir de satisfatoriamente o primeiro momento da série histórica.

Tendo em vista o teste de Levene, onde o objetivo é verificar se a variância dos cenários hidrológicos sintéticos é igual à variância da série histórica, o resultado também pode ser considerado satisfatório para o subsistema Sul, uma vez que o índice de aprovação foi de 94%, e razoável para o subsistema Sudeste/Centro-Oeste que obteve um índice de aprovação de 70%.

O teste de Kolmogorov-Sminov consiste em verificar se os períodos gerados pelo modelo são originários da mesma distribuição dos períodos históricos, ou seja, se ambos possuem distribuições de probabilidades iguais. O índice de aprovação foi de 87% no subsistema Sul e 79% no subsistema Sudeste/Centro-Oeste.

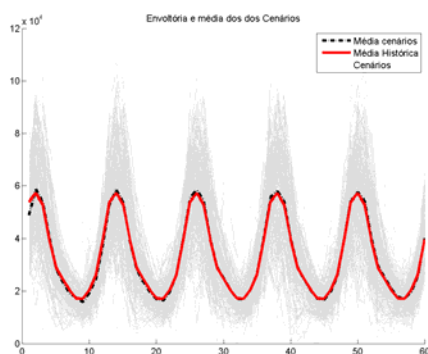


Figura 4 – Envoltória e Médias – Sudeste/Centro-Oeste

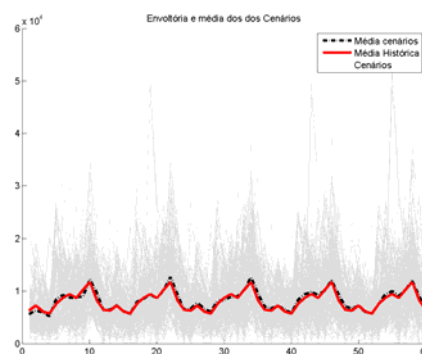


Figura 5 – Envoltória e Médias - Sul

Tabela 4 – Avaliação dos Cenários

	Teste de Média	Teste de Levene	Teste K-S
Sudeste	97%	70%	79%
Sul	99%	94%	87%

Complementando as análises realizadas, neste momento são apresentados os resultados para averiguar a capacidade dos cenários gerados em reproduzir os períodos de seca, ou períodos críticos, observados no histórico.

O objetivo é avaliar se os cenários reproduzem as distribuições de probabilidade de cada variável, comparando-as com a respectiva distribuição histórica. As variáveis empregadas foram previamente definidas e são: comprimento, soma e intensidade de sequência e os testes feitos foram os de Kolmogorov-Sminov para as variáveis soma e intensidade, e o teste Qui-Quadrado para a variável comprimento.

A análise do teste Kolmogorov-Sminov acontece através dos *p-valores* em que esses devem ser superiores ao nível de significância adotado (5%), para que não seja possível rejeitar a hipótese nula. Com relação ao teste Qui-Quadrado, a análise é realizada com base na estatística de teste que deve ser menor do que o valor crítico calculado com base no número de graus de liberdade e no nível de significância.

Na tabela 5, encontram-se os resultados para os testes realizados.

Como pode ser observado, tendo em vista a análise de sequências negativas, as variáveis soma e intensidade de sequência são aderentes ao histórico para o subsistema Sul enquanto que as variáveis comprimento e soma são aderentes no subsistema Sudeste/Centro-Oeste. Por outro lado, a variável comprimento de sequência apresentou diferenças estatísticas entre o histórico e o observado no subsistema Sul e a variável intensidade apresentou diferenças para o subsistema Sudeste/Centro-Oeste.

Com relação análise de sequência positiva, apenas a variável intensidade no subsistema Sudeste/Centro-Oeste apresentou diferenças estatísticas entre o histórico e os cenários gerados. No subsistema Sul, todas as variáveis são estatisticamente iguais ao histórico.

Tabela 5 – Análise de Sequências Negativas

	Comprimento	Soma	Intensidade
	Valor Crítico: 3,84 p-valor mín:0,05 p-valor mín:0,05		
Sudeste	3,50	0,15	0,01
Sul	12,08	0,72	0,45

Tabela 6 – Análise de Sequências Positivas

	Comprimento	Soma	Intensidade
	Valor Crítico: 3,84 p-valor mín:0,05 p-valor mín:0,05		
Sudeste	0,45	0,32	0,02
Sul	0,30	0,99	0,45

4. Conclusão

Este trabalho teve como objetivo o estudo do fenômeno de longa dependência na série de energia natural afluyente do subsistema Sul e Sudeste para que fosse possível ajustar um modelo de geração de séries hidrológicas sintéticas.

O *bootstrap* foi aplicado em vários momentos distintos. No que tange à geração de séries hidrológicas sintéticas, utilizou-se a técnica de computação intensiva para a criação de novas séries. Esta aplicação apresentou bons resultados, entretanto, foi evidenciado o problema de afluentes negativos solucionado por meio da seleção de cenários estritamente positivos. Esta abordagem ocasionou uma elevação da média dos cenários, que foi sanado através de intervalos construídos por *moving blocks*. O método empregado para a seleção se comportou de forma eficiente, uma vez que o mesmo permitiu que cenários com médias muito elevadas fossem descartadas e, desta forma, o problema de elevação das médias resolvido.

Com relação aos três primeiros testes (média, Levene e Kolmogorov-Sminorv), o subsistema Sul apresentou resultados superiores ao subsistema Sudeste em todos os teste.

Na análise de sequências negativas, onde o intuito foi avaliar a capacidade do modelo de criar períodos críticos mais severos que o observado no histórico, os resultados podem ser considerados razoáveis. Foram testadas três variáveis em cada subsistema, e em apenas uma não apresentou aderência entre os cenários e o histórico. No subsistema Sudeste/Centro-Oeste a variável que apresentou diferenças foi a variável intensidade enquanto que no subsistema Sul foi a variável comprimento.

Tendo em vista a análise de sequência positiva, o modelo para o subsistema Sul conseguiu reproduzir as três variáveis analisadas enquanto que o modelo do subsistema Sudeste/Centro-Oeste somente não reproduziu a variável intensidade de sequência.

Isto posto, pode-se afirmar que a metodologia utilizada é capaz de incorporar efeitos de longa dependência e gerar séries sintéticas diferentes do histórico, porém equiprováveis do ponto de vista estatístico.

O modelo do subsistema Sul apresentou resultados superiores em todas as análises realizadas quando comparado com os resultados obtidos para o subsistema Sudeste/Centro-Oeste.

Vale destacar também o pequeno número de parâmetros estimados para construção dos modelos. No caso, são quatro parâmetros para o subsistema Sul e cinco parâmetros para o subsistema Sudeste/Centro-Oeste.

Referências

- EFRON, B.** (1979), Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, v. 7, pp. 1-26.
- GRANGER, C. W. J.; JOYEUX, R.** (1980), An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, v.1, pp.15-29.

- HOSKING, J. R. M.** (1981), Fractional Differencing. *Biometrika*, v.68(1), pp. 165-176.
- OLIVEIRA, F. L. C.** Nova Abordagem para Geração de Cenários de Afluências no Planejamento da Operação Energética de Médio Prazo. *Dissertação de Mestrado*, Departamento de Engenharia Elétrica, PUC-Rio, Brasil, 2010.
- THOMAS, H. A.; FIERING, M.B.** Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. *Design Water Resource Systems*, edited by A.Mass el al.,pp 459-463, Harvard University Press, Cambrigde, Mass., 1962.
- MARCATO, A. L. M.** Representação híbrida de sistemas equivalentes e individualizados para o planejamento da operação de médio prazo de sistemas de potência de grande porte. *Tese de Doutorado*, Departamento de Engenharia Elétrica, PUC-Rio, Brasil, 2002.
- PEIRES, M. S.; SINGH, N.,** (1996). Predictors for Seasonal and Nonseasonal Fracionally Integrated ARIMA Models. *Biometrics*, 38 (6), 741-752.
- PENNA, D. D. J.;** Definição da árvore de cenários de afluências para o planejamento da operação energética de médio prazo. *Tese Doutorado*, Departamento de Engenharia Elétrica, PUC-Rio, 2009.
- PEREIRA, G.A.A;** Modelos de memória longa para geração de cenários hidrológicos sintéticos. *Dissertação de Mestreado*, Departamento de Engenharia Elétrica, PUC-Rio, 2011.
- PORTER-HUDAK, S.** (1990), An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal American Statistical Assoc.* 85, 338-344.
- REISEN, V. A. RODRIGUES, A. L.; PALMA, W.** (2006a), Estimation of seasonal fractionally integrated process. *Computational Statistics e Data Analysis*, 50, 568-582. 2006a.
- REISEN, V. A. RODRIGUES, A. L.; PALMA, W.** (2006a), Estimation seasonal long-memory processes: a Monte Carlo study. *Journal of Statistics Computation and Simulation*, 76 (4), 305-316.
- SOUZA, L. R.;** Implementação de Bootstrap na Estimação do Parâmetro d em Modelos ARFIMA e Simulação Monte Carlo. *Dissertação de Mestrado*, Departamento de Engenharia Elétrica, PUC-Rio, 1997.
- SOUZA, R. C.; CAMARGO, M.E.** *Análise e Previsão de Séries Temporais: Os modelos ARIMA*, Ijuí: SEDIGRAF, 2004.