

GERAÇÃO DE SÉRIES SINTÉTICAS DE AFLUÊNCIAS MENSAS DE INTERESSE ENERGÉTICO

Daniel Henrique Marco Detzel¹; Miriam Rita Moro Mine²; Marcelo Rodrigues Bessa³; Claudio Andres Villegas Vallejos⁴; Eloy Kaviski⁵; Márcio Luís Bloot⁶; Carlos Fernando Bley Carneiro⁷

Resumo --- Um estudo relativo às afluições a sete usinas hidrelétricas da região Sul do Brasil é apresentado. O objetivo principal foi estruturar um modelo estocástico linear para a geração sintética de vazões em escala mensal. Para tanto, as séries históricas de cada usina foram analisadas minuciosamente, partindo-se da verificação da condição de estacionariedade estatística. Optou-se por não trabalhar com modelos periódicos e, dessa maneira, utilizou-se da técnica de dessazonalização das séries. Séries estacionárias e dessazonalizadas foram aplicadas à transformação Box-Cox e submetidas aos procedimentos elementares de identificação, estimação e validação do modelo. Em meio a essas análises, uma formulação parcimoniosa AR(1) foi escolhida e as séries geradas para as sete usinas foram capazes de manter as principais características das observações históricas, além das propriedades residuais inerentes à teoria do modelo.

Abstract --- A study on the inflows to seven hydropower plants in Brazil southern region is presented. The main objective was to design a linear stochastic model for generating synthetic streamflow on a monthly scale. For this, each plant time series were analyzed in detail, starting from statistical stationarity verification. It was chosen to work with non-periodic models and, thus, deseasonalization technique was used for series adjustment. Stationary and deseasonalized series were applied to the Box-Cox transformation and subjected to elementary procedures for identification, estimation and model validation. Through these tests, a parsimonious AR(1) formulation was chosen and the series generated for the seven plants were able to maintain the main historical observations features, and the residual properties inherent in the model theory.

Palavras-Chave – vazões mensais, modelo autorregressivo, série sintética.

- 1) Pesquisador no LACTEC. BR-116, Km 98, Nº 8813 – Centro Politécnico da UFPR, 81531-980 Curitiba. E-mail: daniel@lactec.org.br
- 2) Professora Associada III da UFPR, Centro Politécnico, Jardim das Américas, Curitiba. E-mail: mrmine.dhs@ufpr.br
- 3) Pesquisador no LACTEC. BR-116, Km 98, Nº 8813 – Centro Politécnico da UFPR, 81531-980 Curitiba. E-mail: bessa@lactec.org.br
- 4) Pesquisador no LACTEC. BR-116, Km 98, Nº 8813 – Centro Politécnico da UFPR, 81531-980 Curitiba. E-mail: claudioav@lactec.org.br
- 5) Professor Adjunto da UFPR, Centro Politécnico, Jardim das Américas, Curitiba E-mail: eloy.dhs@ufpr.br
- 6) Pesquisador na COPEL. R. José Izidoro Biazetto, 158 - Santo Inácio, 81200-240 Curitiba. E-mail: mlbloot@copel.com
- 7) Pesquisador na COPEL. R. José Izidoro Biazetto, 158 - Santo Inácio, 81200-240 Curitiba. E-mail: carlos.bley@copel.com

1 - INTRODUÇÃO

Em se tratando de sistemas para o gerenciamento dos recursos hídricos disponíveis em certa região, estudos relativos ao planejamento e operação são de fundamental importância. Sabe-se que o Brasil é um país privilegiado em termos de disponibilidade hídrica para diversos usos fazendo com que o emprego de modelos matemáticos para simular o comportamento das bacias hidrográficas seja muito comum. Para obter as variáveis de interesse aos estudos, a grande maioria dos modelos necessita de séries de registros hidrológicos como dados de entrada. Cita-se como exemplo o caso do planejamento energético nacional: os critérios de suprimento são baseados em índices de risco que são obtidos a partir da simulação da operação energética do sistema. Para estimá-los com incertezas aceitáveis, a série histórica de afluições às usinas não é suficiente, sendo necessário o emprego de uma modelagem probabilística destas vazões. Em muitos outros sistemas hídricos, o planejamento e/ou operação também depende de modelagens desse tipo.

Nesse contexto uma maneira mais eficiente de se utilizar as informações hidrológicas disponíveis é imaginar que a “série histórica” é apenas uma das possíveis realizações de um processo estocástico, ou seja, pode-se imaginar que ela foi sorteada pela “natureza”, segundo um conjunto de leis probabilísticas. Um novo sorteio resultaria em outra série, diferente da histórica, mas igualmente provável.

O presente estudo foi elaborado com vistas ao planejamento da geração hidrelétrica nacional. O panorama atual do parque gerador brasileiro evidencia a importância dos estudos relativos às vazões afluentes às usinas, ao atestar que cerca de 75% da energia gerada provenha de usinas hidrelétricas (MME/EPE, 2010). Aliado a esse fato, destaca-se que grande parte do risco considerado na operação do sistema advém da incerteza quanto ao comportamento hidrológico das vazões afluentes aos reservatórios. Técnicas de modelagem probabilística de vazões, por meio de geração sintética de séries de afluições, são ferramentas extremamente atrativas para a solução desse problema.

Uma questão relevante quando se trabalha com modelagem de vazões é a dependência em série, também chamada de persistência (Kelman, 1987). Diferentemente de outras variáveis hidrológicas (como a precipitação, por exemplo), não se pode adotar a hipótese de aleatoriedade entre os dados de um mesmo rio em intervalos de tempo diferentes. Esse fato faz com que pesquisadores interessados em gerar séries sintéticas de vazões se concentrem em modelos autorregressivos $AR(p)$ ou autorregressivos médias móveis $ARMA(p,q)$. Destacam-se três referências consagradas no tocante a essas formulações, contendo estudos detalhados sobre cada modelo: Box e Jenkins (1976, com edição mais recente em Box *et al.*, 1994), Hipel e McLeod (1994) e Salas *et al.* (1980), este último com aplicações específicas a séries hidrológicas.

Outro fator importante se refere à discretização temporal desejada para as séries. Vários estudos disponíveis na literatura trazem métodos de geração aplicados a escalas anuais, mensais, semanais e diárias. Este texto, em concordância com o propósito do artigo, concentra-se na escala mensal. Modelos com essa discretização temporal podem ser formulados de duas maneiras: geração de vazões anuais com posterior desagregação em mensais ou geração mensal diretamente.

Recentemente, Celeste *et al.* (2005) demonstraram uma aplicação de um modelo de primeira ordem AR(1), gerando séries sintéticas de aflúências para sete reservatórios localizados no sudoeste da Paraíba. Mesmo respeitando as estatísticas básicas do modelo, os autores constataram que as séries tiveram um comportamento completamente aleatório, quando comparadas aos registros históricos. Wang (2008) utilizou um modelo muito semelhante na região de Melbourne, Austrália, obtendo resultados consistentes mesmo contanto com séries históricas não consistidas ou com muitos erros de observação.

Ambos os trabalhos supramencionados não se referem a escalas mensais, contudo as séries geradas podem ser submetidas a técnicas de desagregação. Os modelos de desagregação em hidrologia foram primeiramente sugeridos por Valencia e Shaake (1973). Esta técnica preserva as matrizes de covariância entre as vazões mensais e entre as vazões mensais e anuais das séries histórica e sintética. Entretanto, Mejia e Rousselle (1976) observaram que o modelo possui um revés na representação das correlações interanuais, sugerindo uma alteração na formulação. Uma terceira técnica de desagregação é o chamado método dos cenários hidrológicos, cuja aplicação é significativamente mais simples do que os supracitados e produz resultados tão bons ou até superiores a eles (Groszewicz *et al.*, 1991; Maheepala e Perera, 1996).

A geração direta de aflúências mensais é relativamente mais criteriosa do que anuais, pois a sazonalidade aparece como mais um fator a ser considerado na construção dos modelos. Uma possibilidade é o uso de modelos ARMA(p,q) aplicados à série transformada (ou padronizada). Hipel e McLeod (1994) classificam esse procedimento como *dessazonalização*, no qual todas as características sazonais das séries são retiradas da formulação. Alternativamente, empregam-se modelos periódicos PAR(p) ou PARMA(p,q). Exemplos de modelos periódicos podem ser encontrados em Rasmussen *et al.* (1996), Haltiner e Salas (1988) e Moura e Mendonça (2005).

Outra categoria de modelo autorregressivo, mais abrangente do que a formulação ARMA(p,q), é o modelo ARIMA(p,d,q) (Autorregressivo Integrado Médias Móveis). Esse modelo apresenta um operador diferença d em seu equacionamento que permite representar o comportamento não estacionário das séries hidrológicas. Dentre diversas aplicações desta formulação, destacam-se Lungu e Sefe (1991), Sen (1978) e Mine (1994).

A estruturação de um modelo de geração não é uma tarefa simples e exige muitos cuidados por parte dos pesquisadores. Alguns deles, enumerados por Stedinger e Taylor (1982), podem ser

resumidos: (i) obter registros históricos e outras informações relevantes; (ii) identificar o modelo mais apropriado; (iii) estimar os parâmetros do modelo; (iv) verificar a implementação computacional e (v) validar o modelo em termos das séries geradas. O presente estudo teve por objetivo empregar essas recomendações para montar e validar um modelo de geração de séries sintéticas de aflúências em escala mensal aplicado a sete usinas hidrelétricas da região Sul do país. As séries históricas destas usinas foram estudadas sob diversos aspectos, como verificação da condição de estacionariedade estatística e análise das funções de autocorrelação e autocorrelação parciais para a escolha do modelo autorregressivo mais apropriado. Optou-se por utilizar a técnica de dessazonalização mencionada anteriormente, em detrimento do uso de um modelo periódico.

O artigo está estruturado em quatro seções, sendo a primeira referente a esta introdução. A segunda seção descreve em detalhes a área de estudo e os métodos empregados na estruturação do modelo. A terceira seção apresenta os resultados obtidos com o modelo, acompanhado de alguns comentários. Finalmente, a última seção conclui o trabalho.

2 - MÉTODOS EMPREGADOS E ÁREA DE ESTUDO

A presente seção tem por objetivo expor as diversas técnicas utilizadas no desenvolvimento deste trabalho. A sequência de itens foi estruturada de maneira lógica, condizente com a montagem do modelo estocástico empregado. Dessa maneira, descrever-se-ão os métodos para verificação da estacionariedade estatística das séries, preparação dos dados, escolha do modelo, estimação dos parâmetros e validação dos resultados.

2.1 - Verificação da estacionariedade estatística das séries hidrológicas

Um grande número de modelos hidrológicos existentes que fazem uso de séries históricas para o cálculo de seus parâmetros considera, implicitamente, uma condição de estabilidade natural dos processos envolvidos. Essa condição é conhecida por *estacionariedade* e se refere a um estado de equilíbrio, no qual os momentos estatísticos de uma série são considerados invariantes no tempo (Batista *et al.*, 2009). Ainda, Clarke (2007) afirma que os reflexos da não estacionariedade na produção hidrelétrica são questões de extrema importância e que oferecem grandes desafios aos hidrólogos nos dias de hoje. Como o presente artigo trabalha com séries de aflúências a usinas hidrelétricas, o início do estudo se deu justamente com a verificação da estacionariedade das séries.

Dentre algumas formas de se lidar com a não estacionariedade, podem-se destacar duas: (i) uso de modelos autorregressivos com operador diferença ARIMA(p,d,q) ou (ii) detecção e remoção da não estacionariedade através de pré-processamento dos dados. Ao relevar esses dois métodos, optou-se por seguir o segundo, visto que a formulação ARIMA exige a estimação de uma carga extra de parâmetros e, conseqüentemente, eleva o custo computacional do modelo.

A verificação da estacionariedade se deu através da aplicação de cinco testes estatísticos consolidados na literatura como métodos eficientes para tal fim (Buishand, 1984; Chen e Rao, 2002; Müller *et al.*, 1998). Os testes utilizados podem ser divididos em duas classes: testes paramétricos, feitos sobre as médias amostrais das séries, e não paramétricos, que buscam tendências independentemente de momentos estatísticos. Outra importante diferença entre as duas classes é que, nos testes paramétricos, há a necessidade de se adotar uma distribuição marginal de probabilidades. Para evitar assumir uma distribuição probabilística específica para todas as usinas do estudo, optou-se por aplicar apenas um teste paramétrico: t-Student (Welch, 1947). Os demais testes foram: Cox-Stuart, Wilcoxon e Coeficiente de Correlação de Spearman (Siegel e Castellan Jr., 1988) e Mann-Kendall (ELETROBRÁS, 1987). Todos os testes foram bilaterais com um nível de significância de 5%.

Para séries consideradas estatisticamente não estacionárias, um método expedito de correção foi empregado. Utilizado também em Batista *et al.* (2009), ele parte do princípio de que, em uma série estacionária, é esperado que o traçado gráfico da curva acumulativa de vazões em uma estação fluviométrica possa ser ajustado por uma linha de tendência contínua ao longo de todo o período. Uma mudança na declividade da curva faria com que duas retas de tendência possam ser ajustadas, uma para cada subperíodo. Essa condição, por sua vez, caracterizaria uma série não estacionária que pode ser corrigida de forma simples, através da determinação dos coeficientes angulares das duas retas. Multiplica-se o coeficiente do período mais recente aos dados do período mais antigo, linearizando a tendência.

2.2 - Modelo estocástico para geração das aflúências sintéticas mensais

O processo utilizado para a modelagem das aflúências mensais é do tipo estocástico linear estacionário não periódico. A formulação de um processo desse tipo segue, genericamente, a equação (Box e Jenkins, 1976; Box *et al.*, 1994):

$$z_t = \sum_{i=1}^{\infty} \phi_i z_{t-i} + a_t + \sum_{j=1}^{\infty} \theta_j a_{t-j} \quad (1)$$

onde z_t representa a série temporal, a_t representa a série de resíduos (ou ruídos aleatórios), t o índice temporal ($t = 1, 2, \dots, n$), n o número de elementos da amostra, ϕ o parâmetro autorregressivo e θ o parâmetro de médias móveis. Em particular, quando $\theta = 0$ e $\phi \neq 0$, tem-se um modelo puramente autorregressivo, ou AR(p); quando $\theta \neq 0$ e $\phi = 0$, tem-se um modelo puramente de médias móveis, ou MA(q); quando $\theta \neq 0$ e $\phi \neq 0$, tem-se um modelo misto autorregressivo médias móveis, ou ARMA(p,q).

Os procedimentos descritos nos próximos itens se referem ao método iterativo de seleção do modelo, proposto inicialmente por Box e Jenkins (1976), que consiste em três passos: identificação, estimação e validação. Informações extraídas das séries históricas são utilizadas nesses passos na intenção de estruturar a formulação mais adequada a cada caso. Primeiramente, contudo, são descritos os métodos empregados na preparação das séries hidrológicas para submissão ao modelo.

2.2.1 - Preparação dos dados

A formulação autorregressiva utilizada para a geração das séries sintéticas se fundamenta na distribuição Normal das aflúências. Todavia, como dito anteriormente, não se pode afirmar que as aflúências a todas as usinas seguem uma distribuição probabilística com essas características. Por esse motivo, optou-se por empregar uma transformação numérica, na intenção de normalizar as séries. Uma técnica bastante conhecida para tal tarefa é o método de Box-Cox (Box e Cox, 1964) dado por:

$$w_{t,m} = \begin{cases} \frac{(z_{t,m} + c)^\lambda - 1}{\lambda} & ; \lambda \neq 0 \\ \ln(z_{t,m} + c) & ; \lambda = 0 \end{cases} \quad (2)$$

onde $w_{t,m}$ é a série transformada, $z_{t,m}$ é a série original (estatisticamente estacionária), t é o índice dos elementos da amostra ($t = 1, 2, \dots, n$), m é o índice dos meses do ano e λ e c são os parâmetros da transformação a serem estimados.

Embora se possa ter uma transformação para cada estação sazonal, assume-se um mesmo λ para toda a série, visando diminuir o número de parâmetros a serem estimados (Hipel e McLeod, 1994). No presente estudo o parâmetro c foi excluído da formulação, pois sua presença só é necessária quando se trabalha com dados que incluem valores negativos.

A transformação Box-Cox não é capaz de remover a sazonalidade das séries. Como se pretende trabalhar com um modelo não periódico, é empregado um método de dessazonalização, ou padronização dos dados pela média e desvio padrão, definido por Hipel e McLeod (1994) como:

$$y_{t,m} = \frac{w_{t,m} - \bar{\mu}_m}{\bar{\sigma}_m} \quad (3)$$

onde $y_{t,m}$ é a série dessazonalizada, $w_{t,m}$ é a série transformada conforme (2), $\bar{\mu}_m$ é a média no período m e $\bar{\sigma}_m$ é o desvio padrão no período m .

2.2.2 - Identificação do modelo

A primeira fase da modelagem propriamente dita, identificação do modelo, diz respeito a estudos acerca da classe dos modelos autorregressivos e suas respectivas ordens. Fez-se uso de métodos gráficos, nos quais são grafadas as funções de autocorrelação (FAC) e autocorrelação

parcial (FACP) amostrais das séries estacionárias, transformadas e dessazonalizadas. A partir da comparação dessas funções amostrais com seu comportamento esperado teórico pode-se ter uma primeira aproximação do modelo a ser utilizado. Box *et al.* (1994, p. 187) trazem uma tabela resumindo os comportamentos característicos esperados das funções supramencionadas para modelos AR, MA e ARMA. A mesma referência mostra o equacionamento completo para a obtenção das funções FAC e FACP.

Outro método empregado nesta fase se baseia na determinação do chamado Critério de Informação de Akaike (AIC – Akaike, 1974). Este é um critério matemático, calcado no princípio da parcimônia, ao confrontar as funções de log-verossimilhança com penalidades atreladas ao número de parâmetros do modelo (Box *et al.*, 1994). A combinação das funções e penalidades que resultar no menor AIC indica o modelo mais adequado. Nota-se que, diferentemente do método gráfico, alguns modelos candidatos precisam ser selecionados e ter seus parâmetros estimados para a verificação desse critério. Neste trabalho, foram pré-selecionados cinco modelos: AR(1), AR(2), ARMA(1,1), ARMA(2,1) e ARMA(2,2). Vale lembrar que as afluições possuem forte característica de correlação entre os registros e, por esse motivo, nenhum modelo puramente MA foi testado. Ademais, não foram testados modelos de ordens superiores a dois, pois, segundo Box *et al.* (1994, p. 11), séries temporais estacionárias são representadas apropriadamente com modelos estocásticos lineares com ordens limitadas a dois.

2.2.3 - Estimação dos parâmetros do modelo

Dentre os diversos métodos para estimações dos parâmetros de um modelo estocástico, escolheu-se o Método da Máxima Verossimilhança, por apresentar bons resultados em amostras com tendências assintóticas. Em termos específicos, foi utilizado o Método da Máxima Verossimilhança Condicional que depende (ou está condicionado) de um valor inicial atribuído aos parâmetros para começar o processo iterativo que resulta nos valores finais. Seguiram-se recomendações de Mine (1984) e Salas *et al.* (1980), que separaram o procedimento em duas etapas: (i) estimacão preliminar e (ii) estimacão ótima através do método da Máxima Verossimilhança Condicional propriamente dito. Na primeira etapa, estimaram-se os parâmetros com as informações extraídas diretamente da amostra. Na segunda etapa, buscou-se uma soluçao para a equaçao de log-verossimilhança (Mine, 1984):

$$\ln L(\hat{\phi}, \hat{\theta}, \hat{\sigma}_a^2 / y_t) = -n \cdot \ln \hat{\sigma}_a - \frac{S(\hat{\phi}, \hat{\theta})}{2\hat{\sigma}_a^2} \quad (4)$$

onde $\hat{\phi}$ e $\hat{\theta}$ são os parâmetros a serem estimados, n é o tamanho da amostra, $\hat{\sigma}_a^2$ é a estimativa da variância dos resíduos e $S(\hat{\phi}, \hat{\theta})$ é a chamada função soma dos quadrados dos resíduos. Esta função, em particular, tem importância acentuada no processo de estimacão dos parâmetros do modelo.

Opera-se a equação (1) de modo a isolar o termo dos resíduos a_t ; substituindo-se valores da série histórica nos elementos z_t , os valores numéricos para a função soma dos quadrados dos resíduos são obtidos através da equação:

$$S(\hat{\phi}, \hat{\theta}) = \hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2 + \dots + \hat{a}_n^2 \quad (5)$$

O método iterativo busca uma combinação de parâmetros que maximize a equação (4), ou minimize a função $S(\hat{\phi}, \hat{\theta})$. No presente trabalho, as estimativas preliminares obtidas na primeira etapa foram utilizadas para iniciar o algoritmo de minimização da equação (5). Ressalta-se que ambos os parâmetros necessitam estar entre os limites $-1 \leq \phi \leq 1$ e $-1 \leq \theta \leq 1$ para respeitar as condições de estacionariedade e invertibilidade do modelo (Box *et al.*, 1994; Hipel e McLeod, 1994; Salas *et al.*, 1980).

Uma vez selecionado o modelo e estimados os seus parâmetros, as séries sintéticas foram geradas. Nesta fase, a equação (1), ajustada ao modelo resultante, foi novamente utilizada. A diferença está nos elementos a_t que, de resíduos, passam a ser variáveis aleatórias $a_t \sim N(0, \hat{\sigma}_a^2)$, obtidas com auxílio de um gerador de números aleatórios (métodos para geração de tais números podem ser conferidas em Kaviski, 2006).

2.2.4 - Validação do modelo

De acordo com Haltiner e Salas (1988) a validação do modelo se baseia em três etapas, com crescente nível de rigorosidade: (i) verificação das propriedades dos resíduos obtidos com o modelo ajustado; (ii) preservação das estatísticas de curto termo e (iii) preservação das estatísticas de longo termo.

Os resíduos fornecem boas medidas da qualidade de adequação da formulação autorregressiva aos dados em uso. Para tanto, a série \hat{a}_t precisa ser independente entre si (ruídos brancos), aproximadamente homocedástica (variância constante) e ter distribuição aproximadamente Normal. Esta verificação foi feita aplicando-se testes clássicos: teste de Portmanteau, para independência (Hipel e McLeod, 1994) e teste de Levene, para homocedasticidade (Brown e Forsythe, 1974). No caso da verificação da normalidade, utilizaram-se duas técnicas: teste de Shapiro-Wilk e plotagens Quantil-Quantil (Ferreira, 2008).

A primeira forma de verificação do modelo baseado nas afluições propriamente ditas se relacionou com a verificação das estatísticas de curto termo. Segundo Kelman (1987), as estatísticas de curto termo da série histórica são utilizadas na construção do modelo, fazendo com que a reprodução destes elementos somente confirme que a formulação foi corretamente implementada computacionalmente. As estatísticas de curto termo calculadas nesta etapa foram: médias, desvios padrão, coeficientes de assimetria, autocorrelações de *lags* 1 e 2, vazões máximas e vazões mínimas.

A última etapa da validação do modelo se refere à verificação das estatísticas de longo termo. Os indicadores escolhidos estão relacionados com as épocas de estiagens e acumulação de água em reservatórios. No primeiro caso, as estatísticas foram calculadas conforme a chamada teoria das corridas ou sequências (Haltiner e Salas 1988). A partir de um valor de corte (média de longo termo) foram contados quantos elementos em sequência estavam abaixo dele. Cada conjunto de elementos com essa característica é chamado de *corrida*. Foram determinados o número médio de corridas, a duração máxima de uma corrida, a afluência total média por corrida e a afluência máxima acumulada em uma corrida.

Para os indicadores relacionados à regularização em reservatórios, foi determinado o máximo déficit acumulado (Neira, 2005). Fisicamente, esta grandeza representa o máximo déficit, em termos de afluências, necessário para o sistema suprir a regularização de um reservatório. No presente trabalho, assumiu-se que 80% da afluência média da série seja requerida para tal objetivo. Percebe-se que se trabalha com um sistema sem consideração de falhas, mas, ainda sim, este índice é uma boa ferramenta de análise de desempenho do modelo.

2.2.5 - Descrição da área de estudo

Os métodos descritos nos itens anteriores foram aplicados às afluências de sete usinas hidrelétricas localizadas na região sul do país. Optou-se por dois sistemas de cascata: um pequeno, sobre o rio Uruguai e um mais complexo, sobre o rio Iguaçu, como mostra o esquema da Figura 1.

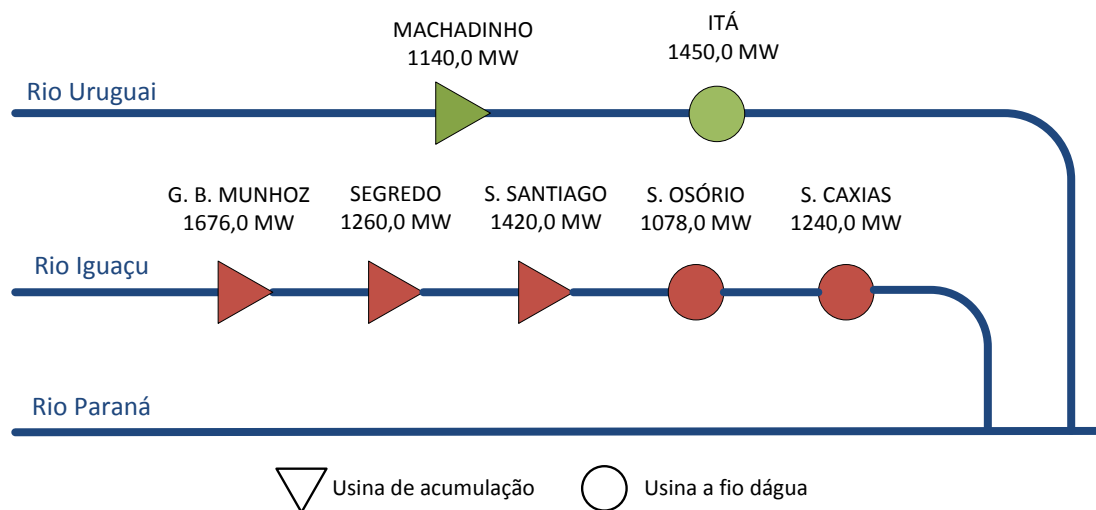


Figura 1 – Usinas Hidrelétricas em estudo

Todas as séries históricas das usinas foram obtidas diretamente do banco de dados do Operador Nacional do Sistema Elétrico (ONS), na forma de registros mensais que totalizaram 77 anos, ou 924 meses, entre as datas de janeiro de 1931 e dezembro de 2007. É importante ressaltar que as séries disponibilizadas pelo ONS se referem às vazões naturais, ou seja, sem contar com efeitos do barramento sobre o rio, além de considerações pertinentes à evaporação e usos secundários. A exceção recai sobre a usina de Segredo que recebe vazões adicionais provenientes

do desvio artificial na usina de Jordão (não contabilizada no estudo). Como esta afluência adicional é aproveitada pelas demais usinas a jusante na cascata, ela foi contabilizada em Segredo.

Outro detalhe importante se refere à qualidade das séries. A análise de consistência das séries hidrológicas é delegada aos agentes responsáveis por cada posto de medição. Métodos específicos para elaboração dessas análises não são determinados pelo ONS, ficando a cargo de cada agente. Ainda assim, as séries provenientes do banco de dados do ONS são consideradas consistidas e não apresentam falhas.

3 - RESULTADOS E DISCUSSÃO

A Figura 2 resume, de forma sequencial, o caminho seguido para a obtenção e validação do modelo estocástico do presente trabalho.

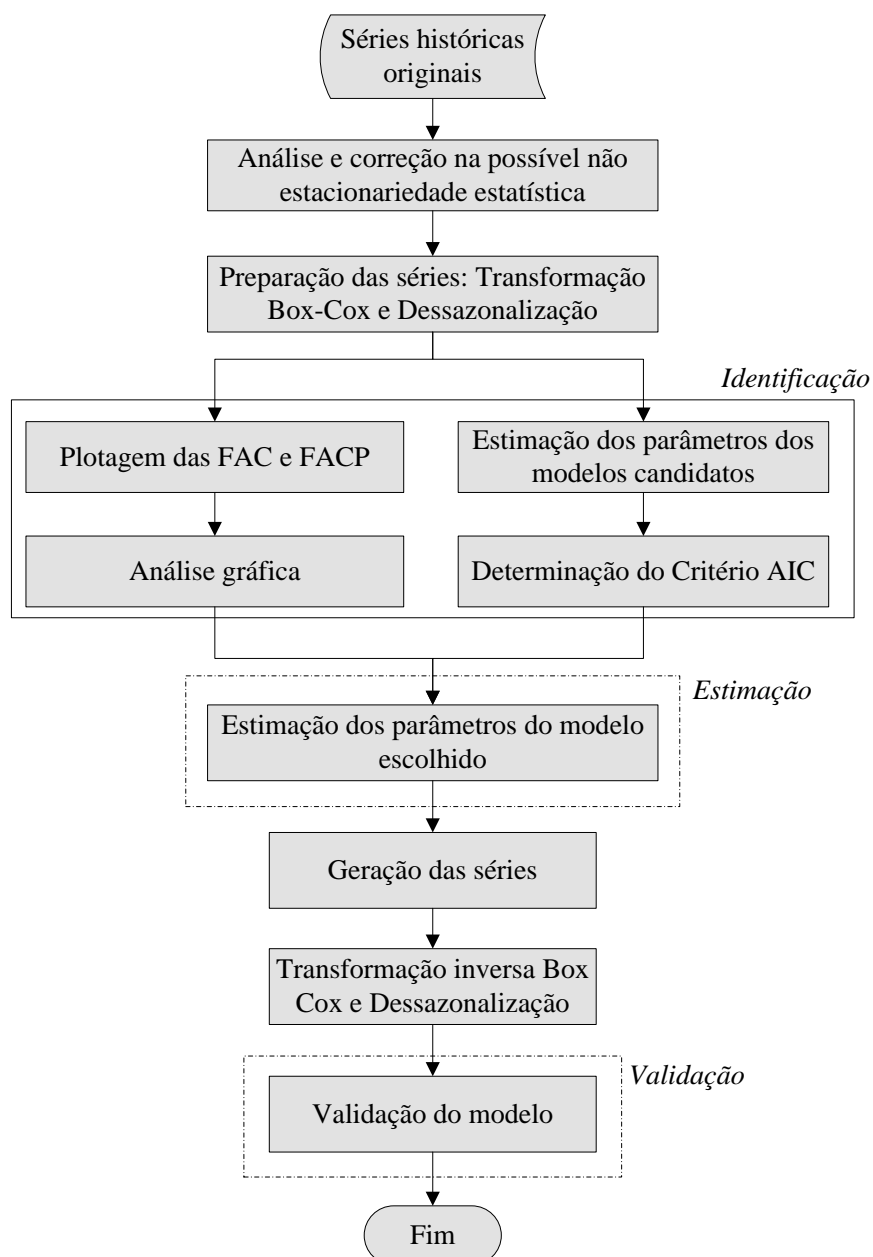


Figura 2 – Resumo do método aplicado

Cada uma das séries de aflúncias foi analisada individualmente, a fim de validar os métodos utilizados, e os resultados são apresentados a seguir.

3.1 - Verificação da estacionariedade estatística das séries hidrológicas

A Tabela 1 mostra os resultados da aplicação dos cinco testes estatísticos às séries de aflúncias das sete usinas do estudo. As abreviações utilizadas são as seguintes: t, para o teste t-Student; CS, para o teste de Cox-Stuart; W, para o teste de Wilcoxon; SP, para o teste do Coeficiente de Correlação de Spearman e MK, para o teste de Mann-Kendall.

Tabela 1 – Resultados dos testes para análise na estacionariedade estatística

Usina	Rio	p-valores				
		t	CS	W	SP	MK
G. B. Munhoz	Iguaçu	0,01	0,08	0,01	0,05	0,05
Segredo	Iguaçu	0,00	0,08	0,00	0,01	0,01
S. Santiago	Iguaçu	0,00	0,08	0,00	0,01	0,01
S. Osório	Iguaçu	0,00	0,08	0,00	0,01	0,01
S. Caxias	Iguaçu	0,00	0,14	0,00	0,01	0,01
Machadinho	Uruguai	0,00	0,04	0,01	0,01	0,01
Itá	Uruguai	0,00	0,14	0,01	0,02	0,02

Optou-se por não exibir os resultados em termos do binário “aceita/rejeita”, mas através dos chamados *p-valores*. Estes índices de probabilidade fornecem uma melhor noção dos resultados de cada teste. Considerando-se que as séries históricas são apenas um acontecimento dentro de um grande mecanismo gerador pode-se dizer, em termos estatísticos, que está se trabalhando com uma amostra retirada de uma população. Assim, os p-valores são interpretados como a probabilidade de se observar, em outra amostra retirada da mesma população, um valor tão grande ou maior do que a observada com a amostra testada. Em meio a esse raciocínio, quanto menores os p-valores maiores as evidências de não-estacionariedade, dado que a formulação de todas as hipóteses nulas assume, *a priori*, a estacionariedade das séries.

Observa-se que todas as sete usinas consideradas possuem forte evidência de não estacionariedade, visto que os p-valores são muito próximos de zero. Assim, todas as séries foram submetidas ao processo de correção mencionado no item 2.1.

3.2 - Identificação do modelo

As funções de autocorrelação e autocorrelação parcial amostrais das aflúncias (estacionárias, transformadas Box-Cox e dessazonalizadas) das sete usinas foram plotadas. A Figura 3 mostra os gráficos obtidos para a usina de S. Santiago, nos quais as linhas vermelhas mostram os valores de cada função e as linhas azuis delimitam o intervalo de confiança abaixo do qual as funções podem

ser consideradas estatisticamente iguais a zero (equivalente a $\pm 2\sigma$). Nota-se que a FAC tem um decaimento exponencial bem definido, enquanto que a FACP sofre um truncamento logo após o primeiro *lag*. Este comportamento é típico de um modelo AR(1), como atesta Box *et al.* (1994). Os gráficos obtidos para as outras usinas do estudo obtiveram comportamento muito semelhante ao mostrado em S. Santiago.

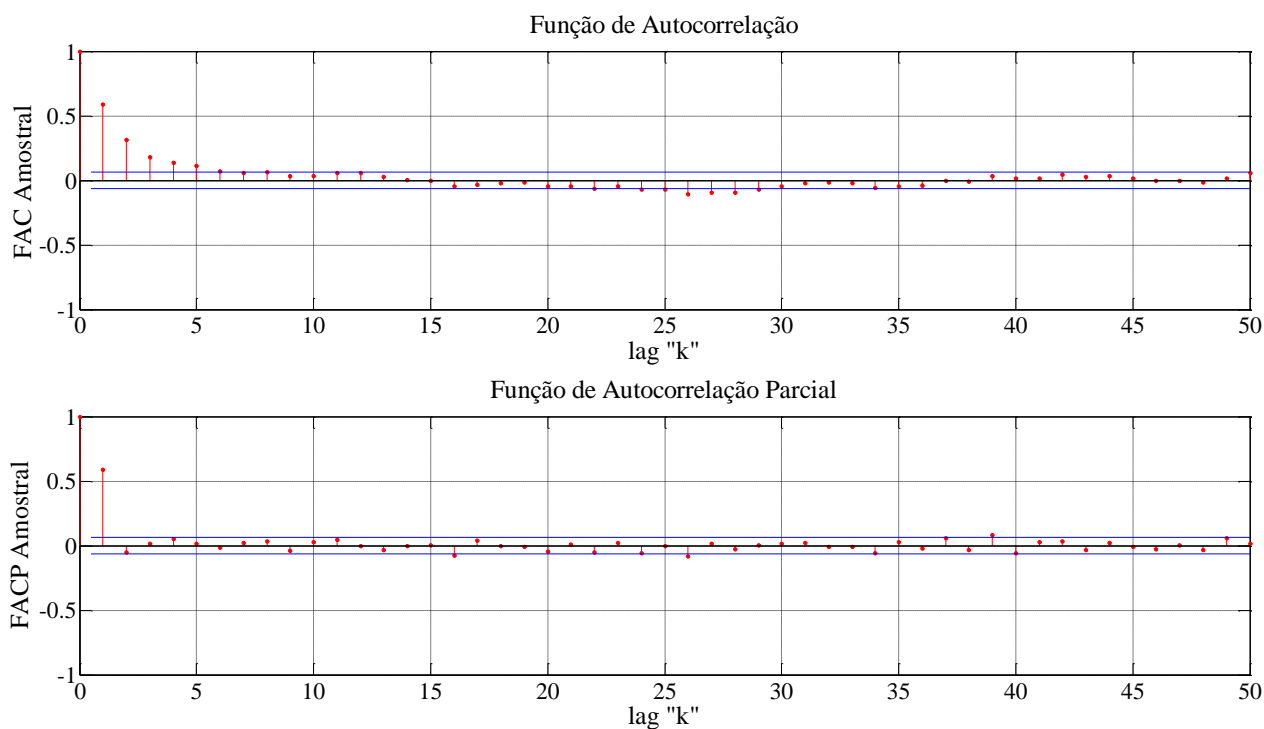


Figura 3 – FAC e FACP para as vazões afluentes de S. Santiago

A segunda forma de identificação do modelo se baseou no Critério de Informação Akaike (AIC). A Tabela 2 mostra os resultados para os cinco modelos pré-selecionados.

Tabela 2 – Resultados do Critério AIC

Usina	AR(1)	AR(2)	ARMA(1,1)	ARMA(2,1)	ARMA(2,2)
G. B. Munhoz	765,90	773,15	770,84	777,56	777,81
Segredo	800,45	809,03	806,32	812,96	813,79
S. Santiago	818,49	828,07	825,46	830,16	833,07
S. Osório	818,00	827,99	825,35	830,05	832,69
S. Caxias	821,03	831,19	828,43	833,30	835,96
Machadinho	658,07	674,27	661,81	677,45	680,65
Itá	642,50	659,26	647,10	662,84	665,86

Nas sete usinas o mínimo AIC se deu para o modelo AR(1), reforçando a indicação dada pelas FAC e FACP amostrais. Dessa forma, este foi o modelo selecionado para a geração das séries sintéticas de afluências nas sete usinas.

3.3 - Estimação dos parâmetros do modelo e geração das séries

Com a identificação de um modelo parcimonioso como o AR(1), a fase de estimação dos parâmetros ficou bastante simplificada, recaindo na determinação de um único coeficiente por série. O estimador de máxima verossimilhança para o modelo AR(1) é fornecido pelo cálculo do coeficiente de correlação *lag* 1 amostral (Box *et al.* 1994). A Tabela 3 mostra os parâmetros para as usinas, bem como a função soma dos quadrados dos resíduos e a equação do modelo resultante.

Tabela 3 – Parâmetros do modelo e equações para geração

Usina	$\hat{\phi}$	$S(\hat{\phi}, \hat{\theta})$	Modelo AR(1)
G. B. Munhoz	0,5736	611,15	$y_t = a_t + 0,5736 y_{t-1}$
Segredo	0,5843	599,82	$y_t = a_t + 0,5843 y_{t-1}$
S. Santiago	0,5897	594,00	$y_t = a_t + 0,5897 y_{t-1}$
S. Osório	0,5896	594,16	$y_t = a_t + 0,5896 y_{t-1}$
S. Caxias	0,5904	593,19	$y_t = a_t + 0,5904 y_{t-1}$
Machadinho	0,5381	647,87	$y_t = a_t + 0,5381 y_{t-1}$
Itá	0,5325	653,36	$y_t = a_t + 0,5325 y_{t-1}$

Para iniciar a geração das séries, é necessário arbitrar um valor inicial para cada usina (representado na equação do modelo por y_{t-1}). O valor arbitrado foi o último elemento da série histórica estacionária, devidamente transformado Box-Cox e dessazonalizado. Para cada usina, foi gerado um conjunto de 1.000 séries sintéticas.

3.4 - Validação das séries

Seguindo a lógica apresentada no item 2.2.4, a primeira verificação se refere às propriedades dos resíduos obtidos com o modelo ajustado. Os resultados obtidos tanto para o teste de Portmanteau quanto para o teste de Levene, apontaram aceites das respectivas hipóteses nulas em todas as usinas consideradas. Entretanto, para o teste de Shapiro-Wilk rejeições foram detectadas. Em uma análise mais aprofundada sobre as causas dessas rejeições, atestou-se que teste de Shapiro-Wilk é muito rigoroso e, por esse motivo, qualquer elemento fora do padrão da amostra resulta na rejeição da hipótese nula (STATA CORP LP, 1991). Em particular, esses desvios acontecem quanto se trabalha com amostras de grande tamanho, como é o caso do presente trabalho. A recomendação passa a ser uma análise gráfica através de histogramas ou plotagens Quantil-Quantil (*Q-Q plots* – Ferreira, 2008).

A Figura 4 mostra gráficos Quantil-Quantil para as usinas G. B. Munhoz e Itá, nas quais todos os 924 elementos de cada amostra estão grafados. Observando a figura, percebe-se que uma reta pode ser ajustada aos pontos, com exceção de alguns poucos elementos nas extremidades. O

comportamento dos $Q-Q$ plots para as demais usinas teve comportamento semelhante. Dessa maneira, conclui-se que a série de resíduos tem distribuição aproximadamente normal, cumprindo o quesito necessário ao bom ajuste do modelo.

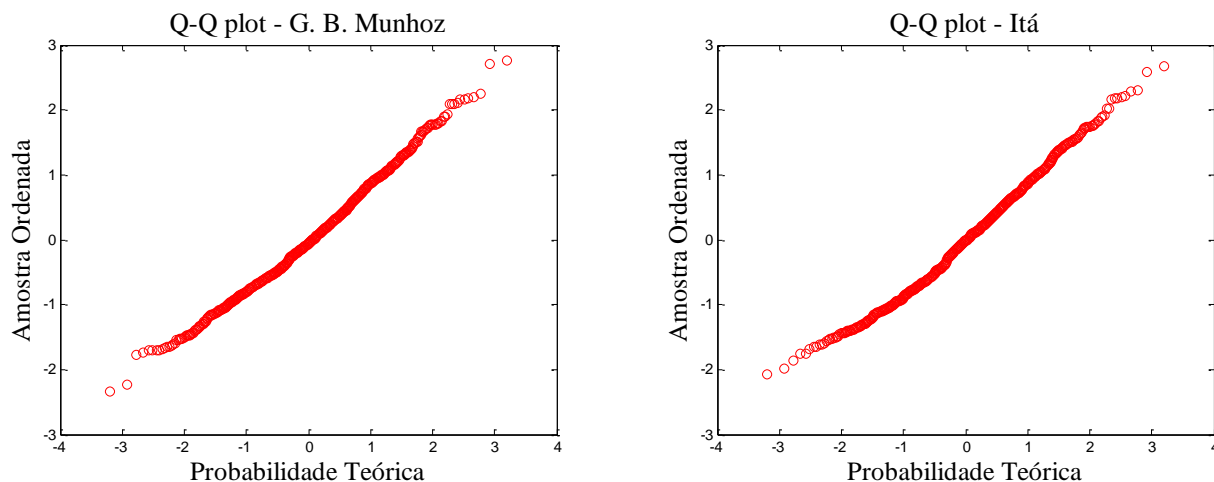


Figura 4 – Q-Q plots para as usinas G. B. Munhoz e Itá

Como primeira forma de validar o modelo em termos de afluências geradas, verificaram-se as estatísticas de curto termo. A Tabela 4 mostra os resultados obtidos; as abreviações utilizadas: MLT – Média; DP – Desvio Padrão; AC 1– Autocorrelação lag 1; AC 2 – Autocorrelação lag 2, SH – Séries Históricas; SS – Séries Sintéticas. Os valores das séries sintéticas se referem à média das séries geradas, com exceção das vazões máximas e mínimas.

Tabela 4 – Resultados das estatísticas de curto termo

Usina	Séries	MLT (m ³ /s)	DP (m ³ /s)	Assimetria	AC 1	AC 2	Máximo (m ³ /s)	Mínimo (m ³ /s)
G. B. Munhoz	SH	722	537	2,3914	0,4923	0,2013	5150	80
	SS	724	563	2,9838	0,4874	0,2353	5791	59
Segredo	SH	958	671	2,0813	0,5044	0,2068	6066	106
	SS	960	703	2,5970	0,5079	0,2492	6770	77
S. Santiago	SH	1135	848	2,4116	0,4923	0,2071	8252	121
	SS	1138	905	3,1300	0,4971	0,2419	9459	100
S. Osório	SH	1188	888	2,3894	0,4924	0,2031	8473	126
	SS	1195	954	3,2373	0,4950	0,2392	10193	106
S. Caxias	SH	1524	1135	2,3530	0,4899	0,2067	10798	191
	SS	1531	1213	3,1239	0,5000	0,2376	12659	140
Machadinho	SH	814	654	2,0152	0,4821	0,2151	5925	56
	SS	819	699	2,6845	0,4964	0,2318	6574	40
Itá	SH	1149	941	2,0017	0,4850	0,2264	8292	62
	SS	1157	1012	2,7789	0,4911	0,2274	9644	64

Na análise dos resultados, percebe-se que as estatísticas de curto termo foram reproduzidas de forma apropriada. O único valor a chamar atenção se refere ao coeficiente de assimetria das séries sintéticas. Em todas as usinas consideradas, este índice foi superestimado. Por outro lado, pode-se destacar o bom desempenho do modelo em reproduzir as autocorrelações, índice de grande importância para séries de afluências mensais.

A terceira e última etapa da validação do modelo se refere à verificação das estatísticas de longo termo. Para a verificação das corridas, definiu-se o valor de corte como sendo a média de longo termo de cada série histórica estacionária (valor igual à MLT, expressa na Tabela 4). Além do número total de corridas, foram calculadas as durações médias (\bar{l}) e máximas ($l_{máx}$), bem como as afluências acumuladas médias (\bar{S}) e máximas ($S_{máx}$) de cada corrida, para cada usina. No caso dos déficits acumulados, determinaram-se os valores em função das afluências médias e máximas para regularizar 80% da média mensal. Todos os resultados estão resumidos na Tabela 5, ressaltando-se que os valores das séries sintéticas se referem à média das séries geradas, excluindo-se os índices que exprimem valores máximos.

Tabela 5 – Resultados das estatísticas de longo termo

Usina	Séries	# Corridas	\bar{l}	$l_{máx}$	\bar{S} (m ³ /s)	$S_{máx}$ (m ³ /s)	Déficit Médio (m ³ /s)	Máx. Déficit Acum. (m ³ /s)
G. B. Munhoz	SH	100	5	19	2187	7443	6089	1162
	SS	99	5	22	2145	8204	7867	1427
Segredo	SH	96	5	19	2962	9887	8187	1487
	SS	96	6	25	3482	13591	9869	1734
S. Santiago	SH	100	5	19	3390	11712	10208	1908
	SS	97	6	23	3604	13883	12922	2376
S. Osório	SH	99	5	21	3554	15269	10764	1976
	SS	96	6	26	4254	16530	13392	2467
S. Caxias	SH	101	5	31	4713	24580	14389	2459
	SS	97	6	24	5066	19503	17047	3129
Machadinho	SH	94	6	27	2433	7964	10432	2075
	SS	98	6	22	2472	8879	9385	1779
Itá	SH	93	6	27	3435	11803	14794	2945
	SS	98	6	22	3513	12700	13705	2615

Analisando somente os dados históricos, nota-se que a estrutura das duas cascatas ficou bem representada através da evolução do número de corridas e seus respectivos índices, no sentido montante - jusante. Ademais, percebe-se a semelhança do regime hidrológico nas duas cascatas consideradas. O modelo estruturado, ao reproduzir com boa precisão essas estatísticas de longo termo, mostrou-se apropriado para as usinas consideradas.

De um modo geral, os resultados menos precisos observados, tanto para as estatísticas de curto quanto para de longo termo, ocorreram para índices máximos (especificamente afluência máxima, número e comprimento máximo de corridas). Esse fato se justifica pela variabilidade amostral esperada quando se trabalha com grandes conjuntos de séries sintéticas e não se mostra preocupante.

Outro revés apresentado pelo modelo foi em relação aos coeficientes de assimetria das séries geradas. Sabe-se que a assimetria de variáveis hidrológicas é um índice importante, diretamente relacionado com a distribuição marginal de probabilidades do processo. Sobre esse assunto, o leitor é referido ao estudo de Lettenmaier e Burges (1977) que buscaram reproduzir o coeficiente de assimetria de forma adequada através do ajuste das distribuições log-normal a três parâmetros e gama a três parâmetros, reconhecidas como bons modelos probabilísticos para séries com forte assimetria. Após ensaios detalhados sobre essas duas distribuições, os autores concluíram não haver vantagens operacionais significativas em usar essas técnicas. Ao invés disso, eles recomendaram dar prioridade à estrutura de persistência das séries. Relevando as conclusões desses autores, percebe-se que o presente trabalho obteve sucesso na estruturação e validação do modelo AR(1) para as usinas em estudo.

4 - CONCLUSÕES

Um estudo extensivo das afluências de sete usinas hidrelétrica localizadas no Sul do Brasil foi apresentado. O objetivo principal foi estruturar um modelo estocástico para geração de séries sintéticas de vazões, em escala mensal. Para tanto, as propriedades amostrais das séries históricas com 924 meses (cada) foram investigadas, na intenção de remover a não estacionariedade estatística e identificar o modelo estocástico linear mais apropriado.

Após essas averiguações, optou-se por uma formulação autorregressiva de primeira ordem não periódica, ou AR(1). Este modelo foi aplicado a séries estacionárias, dessazonalizadas, e normalizadas através da transformação Box-Cox. Apesar de parcimonioso e simples, ele foi capaz de reproduzir com sucesso as diversas características históricas observadas, para as sete usinas. Ademais, considerações teóricas aproximadas relativas aos resíduos do modelo foram também respeitadas.

Com os resultados mostrados, a estruturação de um modelo multivariado fica facilitada, principalmente porque se constatou ser possível modelar as vazões das usinas em uma cascata com um modelo de mesma ordem. Se não for essa a opção, o modelo univariado mostrado neste estudo é apropriado para simulações diversas individuais a cada usina.

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.

BIBLIOGRAFIA

AKAIKE, H. (1974). “*A new look at the statistical model identification*”. IEEE Transactions on Automatic Control 19(6), pp. 716-723.

BATISTA, A. L.; FREITAS JR., S. A. de; DETZEL, D. H. M.; MINE, M. R. M.; FILL, H. D. O. A.; FERNANDES, C.; KAVISKI, E. (2009). “*Verificação da estacionariedade de séries hidrológicas no Sul-Sudeste do Brasil*” in Anais do XVI Simpósio Brasileiro de Recursos Hídricos, Campo Grande, nov. 2009, CD-ROM.

BOX, G. E. P. COX, D. R. (1964). “*An analysis of transformations*”. Journal of the Royal Statistical Society, series B, 26, pp. 211-243.

BOX, G. P.; JENKINS, G. M. (1976) *Time Series Analysis: Forecasting and Control*. Holden-Day, Oakland, 575 p.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. (1994) *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey, 598 p.

BROWN, M. B.; FORSYTHE, A. B. (1974). “*Robust tests for the equality of variances*”. Journal of the American Statistical Association 69(346), 364-367.

BUIHAND, T. A. (1984). “*Tests for detecting a shift in the mean of hydrological time series*”. Journal of Hydrology 73, pp. 51-69.

CELESTE, A. B.; CURI, W. F.; CURI, R. C. (2005). “*Modelo Estocástico Auto-Regressivo Multivariado de Primeira Ordem (AR(1)) para Geração de Séries Sintéticas de Vazões Anuais – Uma aplicação na bacia do Rio Piancó-PB*” in Anais do XVI Simpósio Brasileiro de Recursos Hídricos, João Pessoa, nov. 2005, CD-ROM.

CHEN, H-L.; RAO, A. R. (2002). “*Testing hydrologic time series for stationarity*”. Journal of Hydrologic Engineering 7(2), pp. 129-136.

CLARKE, R. T. (2007). “*Hydrological prediction in a non-stationary world*”. Hydrology and Earth System Sciences 11(1), pp. 408-414.

ELETROBRÁS (1987). “*Guia para cálculo de cheia de projeto de vertedores*”. Rio de Janeiro, pp. 38.

FERREIRA, D. F. (2008). “*Distribuições Multivariadas*”, in: *Estatística Multivariada*. UFLA, Lavras, MG, pp. 107-171.

- GROSZEWICZ, R. C.; KAVISKI, E.; ILLICH, I.; MACHADO, J. F. P. (1991). “*Avaliação de três métodos de desagregação de afluências*” in Anais do IX Simpósio Brasileiro de Recursos Hídricos, Rio de Janeiro, nov. 1991, pp. 437-446.
- HALTINER, J. P.; SALAS, J. D. (1988). “*Development and testing of a multivariate, seasonal ARMA(1,1) model*”. Journal of Hydrology 104, pp. 247-272.
- HIPEL, K. W.; McLEOD, A. I. *Time Series Modelling of Water Resources and Environmental Systems*. Disponível em: <<http://www.stats.uwo.ca/faculty/aim/1994Book/>>. Acesso em 27/04/2011.
- KAVISKI, E. (2006). “*Solução de problemas de fenômenos de transporte pelo método de Monte Carlo*”. Tese (doutorado) – Universidade Federal do Paraná, Setor de Ciências Exatas, Curitiba, 330f. Disponível em: <<http://hdl.handle.net/1884/7839>>. Acesso em: 28/04/2011.
- KELMAN, J. (1987). “*Modelos estocásticos no gerenciamento de recursos hídricos*”, in: *Modelos para Gerenciamento de Recursos Hídricos I*. Barth, et al. ABRH, ed. Nobel/ABRH, São Paulo – SP, pp. 327-424.
- LETTENMAIER, D. P.; BURGESS, S. J. (1977). “*An operational approach to preserving skew in hydrologic model of long-term persistence*”. Water Resources Research 13(2), pp. 281-290.
- LUNGU, E. M.; SEFE, F. T. K. (1991). “*Stochastic analysis of monthly streamflows*”. Journal of Hydrology 126, pp. 171-182.
- MAHEEPALA, S.; PERERA, B. J. C. (1996). “*Monthly hydrologic data generation by disaggregation*”. Journal of Hydrology 178, pp. 277-291.
- MEJIA, J. M., ROUSSELLE, J. (1976). “*Disaggregation models in hydrology revisited*”. Water Resources Research 12(2), pp. 185-186.
- MINISTÉRIO DE MINAS E ENERGIA; EMPRESA DE PESQUISA ENERGÉTICA (2010). “*Características Gerais*”, in *Plano Decenal de Expansão de Energia 2019*. Ed. MME/EPE, Brasília – DF, pp. 21-36. Disponível em: <<http://www.epe.gov.br/PDEE/Forms/EPEEstudo.aspx>>. Acesso em: 27/04/2011.
- MINE, M. R. M. (1984). “*Modelos estocásticos lineares para previsão de cheias em tempo real*”. Dissertação (mestrado) – Universidade de São Paulo, Escola Politécnica, São Paulo, 127 f.
- MOURA, K. L.; MENDONÇA, A. S. F. (2005). “*Modelagem Estocástica de Vazões Mensais para Bacias Hidrográficas Brasileiras*” in Anais do XVI Simpósio Brasileiro de recursos Hídricos, João Pessoa, nov. 2005, CD-ROM.
- MÜLLER, I. I.; KRÜGER, C. M.; KAVISKI, E. (1998). “*Análise de estacionariedade de séries hidrológicas na bacia incremental de Itaipu*”. Revista Brasileira de Recursos Hídricos 3(4), pp. 51-71.
- NEIRA, Katia L. (2005). “*Curvas de Regularização para Reservatórios Parcialmente Cheios e Confiabilidade Constante*”. Dissertação (mestrado) – Universidade Federal do Paraná, Setor de Tecnologia. Curitiba, 164 f.
- RASMUSSEN, P. F.; SALAS, J. D.; FAGHERAZZI, L.; RASSAM, J-C. (1996). “*Estimation and validation of contemporaneous PARMA modelos for streamflow simulation*”. Water Resources Research 32(10), pp. 3151-3160.

- SALAS, J. D.; DELLEUR, J. W.; YEVJEVICH, V.; LANE, W. L. (1980). *Applied Modeling of Hydrologic Time Series*. Water Resources Publications, Littleton, 484 p.
- SEN, Z. (1978). “*A mathematical model of monthly flow sequences*”. *Hydrological Sciences* 23(2), pp. 223-229.
- SIEGEL, S.; CASTELLAN Jr., N. (1988). “*Nonparametric statistics for the behavioral sciences*”. McGraw-Hill, New York, 401 p.
- STATA CORP LP. (1991). “*Stata Technical Bulletin 3*”. Disponível em: <<http://statapress.com/journals/stbcontents/stb3.pdf>>. Acesso em: 01/04/2011.
- STEDINGER, J. R.; TAYLOR, R. T. (1982). “*Synthetic Streamflow Generation: 1. Model Verification and Validation*”. *Water Resources Research* 18(4), pp. 909-918.
- VALENCIA, D. R.; SCHAAKE, J. C. (1973). “*Disaggregation process in stochastic hydrology*”. *Water Resources Research* 9(3), pp. 580-585.
- WANG, Q. J. (2008). “*A Bayesian method for multi-site stochastic data generation: Dealing with non-concurrent and missing data, variable transformation and parameter uncertainty*”. *Environmental Modelling & Software* 23, p. 412-421.
- WELCH, B. L. (1947). “*The generalization of Student's problem when several different population variances are involved*”. *Biometrika* 34(1-2), pp. 28-35.