

Modelos de Inteligência Computacional para Geração de Séries Sintéticas de Vazões Médias Mensais

Ricardo Menezes Salgado

Universidade Federal de Alfenas, Minas Gerais, Brasil

ricardomenezes@ieee.org

Ivette Luna, Rosangela Ballini, Secundino Soares

Universidade Estadual de Campinas, São Paulo, Brasil

{ivette,ballini}@eco.unicamp.br, dino@cose.fee.unicamp.br

Donato da Silva Filho

EDP Bandeirante, São Paulo, Brasil

donato.filho@edpbr.com.br

Resumo – O estudo do comportamento das vazões afluentes desempenha um papel importante no planejamento e na gestão dos recursos hídricos e energéticos. Entretanto, devido à disponibilidade limitada de dados de vazão histórica, a geração de séries sintéticas de vazões médias mensais têm sido amplamente utilizada na simulação de processos hídricos. Este trabalho propõe um procedimento para a geração de séries sintéticas com base em duas abordagens da inteligência computacional: um sistema baseado em regras nebulosas e um modelo de redes neurais multicamadas. Resultados empíricos sobre as séries de vazões mensais associadas a dois postos de vazões da cascata do Rio Grande são apresentados, sendo os resultados obtidos por ambos os modelos confrontados e analisados. Os resultados fornecem evidências que apontam que os modelos utilizados são uma alternativa adequada para a geração de séries sintéticas de vazões mensais e que reproduzem as principais características estatísticas das séries observadas.

Palavras-chave – Séries Sintéticas, Sistema de Inferência Nebulosa, Redes Neurais Artificiais, Hidrologia.

Abstract – The study of behavior of water inflow performs an important role in planning and management of water resources and energy. However, due to limited availability of historical flow data, the generation of monthly inflow synthetic series have been widely used in the water simulation process. This paper proposes a procedure for generating monthly inflow synthetic series using two models based on computational intelligence: a multilayer neural networks (MLP) and a based on fuzzy rules (FIS). Such models will be used to map the hydrological process and generate synthetic sequences of monthly inflow. Empirical tests on the series of monthly inflows associated with different stations in the Grande River are presented and the results obtained by both models confronted and analysed. The results provide evidence that points to the approach proposed as an adequate alternative for the generation of synthetic series of monthly inflow, reproducing the main statistical characteristics of observed series.

Keywords – Synthetic Time Series, Fuzzy Inference System, Artificial Neural Networks, Hydrology.

1. INTRODUÇÃO

Dentre os inúmeros recursos naturais disponíveis, a água destaca-se como um elemento primordial em diversos aspectos. O conhecimento a respeito da disponibilidade de água é essencial devido às suas várias aplicações entre as quais pode-se destacar: abastecimento, sistemas de irrigação, navegação, pesca, recreação, transporte, geração de energia elétrica, entre outras. No Brasil a geração de energia elétrica é predominantemente hidráulica, sendo que a maior parte da capacidade instalada do país é composta por usinas hidroelétricas distribuídas em doze bacias hidrográficas localizadas em diferentes regiões [1].

A produção energética de um sistema hidroelétrico depende da série de vazões afluentes às diversas usinas. Por exemplo, quando se adota um aproveitamento hídrico, a série de vazões no local é usualmente utilizada como dado de entrada para um modelo de simulação e/ou otimização [2].

Em geral, decisões tomadas no gerenciamento de sistemas de recursos hídricos são fundamentadas na análise de registros históricos. A partir de dados observados, são utilizados modelos estocásticos que traduzam a estrutura de correlação temporal da série, permitindo assim a previsão de cenários futuros. De um modo geral, análises estatísticas relacionadas às vazões encontram limitações relacionadas ao número de observações disponíveis, que em sua maioria são insuficientes. Uma alternativa para superar tais limitações é a geração de séries sintéticas.

De acordo com [3], uma série sintética é o resultado da aplicação de um método específico em um conjunto de dados observados, objetivando gerar sequências equiprováveis e que preservem as mesmas características estocásticas dos dados originais.

Na literatura existem diversos modelos para geração de séries sintéticas. As técnicas mais difundidas baseiam-se nos modelos estocásticos autorregressivos e em modelos de simulação baseados nas técnicas de Monte Carlo ([4], [5], [6], [7] e [8]). Alguns

trabalhos recentes utilizam técnicas baseadas em redes neurais artificiais e técnicas de *bootstrap* para geração de sequências sintéticas [9].

Neste artigo é proposto a aplicação de um sistema baseado em regras fuzzy (FIS) e combinado com a técnica de *bootstrap* para a geração de séries sintéticas. A estrutura do modelo fuzzy é determinada pelo algoritmo de agrupamento não supervisionado denominado de *subtractive clustering* [10]. Ou seja, o número de regras fuzzy e os parâmetros das funções de pertinência são obtidos pelo algoritmo de agrupamento. Os consequentes das regras são do tipo *singleton*. Dessa forma, o modelo fuzzy é ajustado com base no histórico de vazões e os resíduos são obtidos. Empregando a técnica de *bootstrap* sob esses resíduos obtêm-se a série sintética.

Além do modelo FIS, neste artigo também é empregado o modelo de redes neurais multi-camadas (MLP) com algoritmo de retro-propagação do erro para a fase de identificação e estimação do processo estocástico. Assim como no FIS, a partir dos resíduos obtidos pela MLP, combina-se a técnica de *bootstrap* para a geração da série sintética.

Neste artigo, os modelos FIS e MLP são aplicados para a geração de séries sintéticas mensais de dois postos localizados na bacia do rio Grande, região Sudeste do Brasil. A partir da análise da assimetria, curtose, desvio padrão e média das séries geradas verificou-se que ambos modelos foram capazes de captar e reproduzir o comportamento das séries históricas de vazões para os dois postos analisados.

Após esta breve introdução, o artigo está organizado da seguinte forma: na seção 2 é apresentado sucintamente o modelo baseado em regras fuzzy FIS. A descrição da metodologia proposta para a geração das séries sintéticas é apresentada na seção 3. Na seção 4 são apresentados os resultados tanto para o modelo FIS como para a rede MLP. Finalmente, a seção 5 é apresentada a conclusão e trabalhos futuros para a geração de séries sintéticas.

2 MODELO BASEADO EM REGRAS FUZZY

A estrutura geral do sistema baseado em regras fuzzy empregada neste artigo, FIS (do termo em inglês *Fuzzy Inference System*), é composta por um conjunto de M regras fuzzy de primeira ordem, do tipo Takagi-Sugeno (TS) [11] (Figura 1).

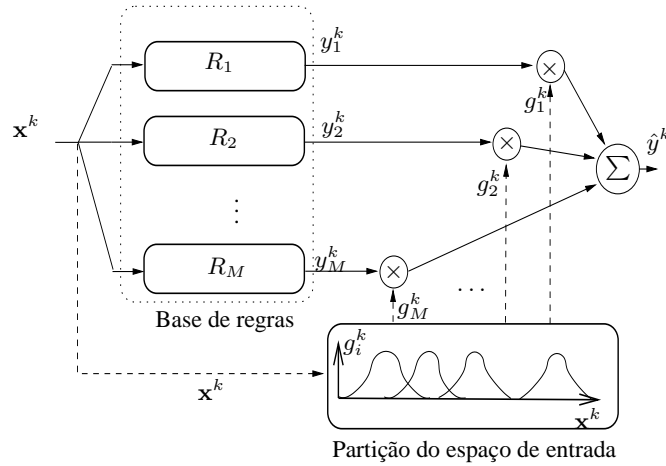


Figura 1: Estrutura geral do FIS.

Seja $\mathbf{x}^k = [x_1^k, x_2^k, \dots, x_p^k] \in \mathbb{R}^p$ o vetor de entrada no instante k , $k \in \mathbb{Z}_0^+$; $\hat{y}^k \in \mathbb{R}$ a saída do modelo que representa em termos de previsão, a estimativa um passo à frente da série em estudo, dada a entrada \mathbf{x}^k .

O espaço de entrada, descrito por $\mathbf{x}^k \in \mathbb{R}^p$ é particionado em M sub-regiões, sendo cada sub-região representada por uma regra fuzzy. Os antecedentes de cada regra fuzzy R_i (do tipo **Se-Então**), são definidos pelos respectivos centros $\mathbf{c}_i \in \mathbb{R}^p$ e matrizes de covariância $\mathbf{V}_i|_{p \times p}$. Os consequentes são descritos por modelos locais lineares com saídas denotadas por y_i^k , $i = 1, \dots, M$ e definidas de acordo com:

$$y_i^k = \phi^k \times \theta_i^T \quad (1)$$

na qual $\phi^k = [1 \ x_1^k \ x_2^k \ \dots \ x_p^k]$; $\theta_i = [\theta_{i0} \ \theta_{i1} \ \dots \ \theta_{ip}]$ é o vetor de coeficientes do modelo local para a i -th regra fuzzy.

A cada padrão de entrada associa-se um grau de pertinência referente às sub-regiões do espaço de entrada. Os graus de pertinência são calculados por funções de pertinência do tipo gaussianas $g_i(\mathbf{x}^k)$, dadas pela seguinte equação:

$$g_i(\mathbf{x}^k) = g_i^k = \frac{\alpha_i \cdot P[i | \mathbf{x}^k]}{\sum_{q=1}^M \alpha_q \cdot P[q | \mathbf{x}^k]} \quad (2)$$

sendo que α_i são coeficientes positivos que satisfazem a condição de unicidade $\sum_{i=1}^M \alpha_i = 1$, e $P[i | \mathbf{x}^k]$ é definida de acordo com:

$$P[i | \mathbf{x}^k] = \frac{1}{(2\pi)^{p/2} \det(\mathbf{V}_i)^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^k - \mathbf{c}_i) \mathbf{V}_i^{-1} (\mathbf{x}^k - \mathbf{c}_i)^T \right\} \quad (3)$$

em que $\det(\cdot)$ representa o operador determinante. Assim, a saída do sistema FIS $y(k) = \hat{y}^k$, é calculada como uma combinação não linear dos consequentes y_i^k ponderados pelos seus respectivos graus de pertinência g_i^k , ou seja:

$$\hat{y}(\mathbf{x}^k) = \hat{y}^k = \sum_{i=1}^M g_i^k y_i^k \quad (4)$$

Neste trabalho a estrutura inicial do modelo FIS é fornecida pelo algoritmo de agrupamento não supervisionado *Subtractive Clustering* (SC) [10]. Ou seja, este algoritmo fornece o número de regras fuzzy assim como os valores modais e de dispersão das funções de pertinência. Vale destacar que esta estrutura é determinada a partir do conjunto de dados históricos de vazões.

3 GERAÇÃO DAS SÉRIES SINTÉTICAS

As séries de vazões apresentam um comportamento periódico, ou seja, são séries sazonais, com período aproximado de doze meses. Esse tipo de comportamento representa um caso particular de não estacionariedade da série. De forma geral, uma série temporal y_t com não estacionariedade sazonal pode ser representada por dois componentes, um sazonal s_t e outro estacionário \tilde{z}_t :

$$y_t = s_t + \tilde{z}_t + a_t \quad (5)$$

sendo a_t um componente aleatório com média zero e variância constante σ_a^2 .

Para a geração das vazões sintéticas mensais, são determinados modelos mensais, ou seja, doze modelos independentemente. Os modelos são ajustados com o objetivo de se estimar o componente estacionário.

O componente aleatório é selecionado a partir do conjunto de resíduos estimados ($e_{m,t}$) dado pela diferença entre a vazão observada do mês m e ano t , $y_{m,t}$, e o valor estimado $\hat{y}_{m,t}$ pelo modelo mensal, sobre o conjunto de dados utilizados durante o ajuste. Formalmente,

$$\begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{12,t} \end{bmatrix} = \begin{bmatrix} \tilde{z}_{1,t} \\ \tilde{z}_{2,t} \\ \vdots \\ \tilde{z}_{12,t} \end{bmatrix} - \begin{bmatrix} \hat{y}_{1,t} \\ \hat{y}_{2,t} \\ \vdots \\ \hat{y}_{12,t} \end{bmatrix}$$

sendo que $e_{m,t}$ a variável aleatória independente entre si, de média zero e variância constante para $m = 1 \dots, 12$. O dado $\hat{y}_{m,t}$ pode ser estimado pelo modelo FIS (ou MLP). Como o componente aleatório é independente, pode-se sortear pontos das amostras de cada componente de $e_{m,t}$.

3.1 Pré-processamento dos dados

A assimetria na distribuição dos dados de vazões é uma característica das séries em estudo que não pode ser desprezada. Existem picos de vazão e períodos de seca que fogem do comportamento médio das vazões ao longo do histórico. Embora os modelos descritos tenham se mostrado adequados em termos de previsão um passo à frente considerando uma distribuição simétrica nos dados (como a normal, por exemplo), este desempenho não se mantém quando os mesmos modelos são utilizados para fins de geração de séries sintéticas, pois uma distribuição simétrica nos impediria de replicar os valores extremos da série. Assim, surge a necessidade de utilizar modelos livres de qualquer pressuposto sobre a distribuição das séries para fins de geração, como é o caso do modelo FIS e da rede neural MLP.

Para tornar a série de vazão estacionária, é necessário remover o componente sazonal. Para isso realiza-se a seguinte transformação nos dados:

$$\tilde{z}_{m,t} = \frac{y_{m,t} - \hat{\mu}_m}{\hat{\sigma}_m} \quad (6)$$

na qual, $\hat{\mu}_m$ e $\hat{\sigma}_m$ são a média e desvio padrão estimados para cada mês m , respectivamente. Após toma-se os valores máximos e mínimos mensais da variável $\tilde{z}_{m,t}$ e padroniza-se os dados no intervalo $[0, 1]$.

3.2 Identificação dos modelos

Um passo fundamental no momento de ajustar um modelo de previsão e/ou geração de vazões é a escolha das entradas. A relação existente entre o par (entrada, saída desejada) é responsável por fornecer ao modelo as informações necessárias para prever ou gerar a vazão mensal. Erros neste processo ocasionam redução da capacidade de generalização do modelo podendo implicar em perda na eficiência do gerador de séries sintéticas.

Neste trabalho, para o modelo FIS o número de atrasos foi escolhido pelo critério Bayesian Information Criterium (BIC), proposto por [12]. No caso da rede MLP, o número de atrasos foi determinado pela minimização do erro médio quadrático

(MSE) [13]. Assim, para ajustar os modelos para um determinado mês i foram calculados os critérios BIC ou MSE variando o número de atrasos $p = 1, \dots, 6$. O valor de p que apresentou menor índice foi escolhido para representar as informações dos padrões de vazões. O critério MSE também foi empregado para a escolha do número de neurônios na camada intermediária, taxa de aprendizado e número de épocas.

3.3 Geração da Série Sintética

De acordo com a equação (5), a série de vazão pode ser representada pelos seguintes componentes: sazonal, estacionário e aleatório. Após a transformação nos dados, dada pela equação (6), um modelo FIS ou MLP é ajustado à série de vazões estacionária.

No processo de geração da série sintética, para determinado vetor de entrada o modelo FIS/MLP estima a vazão para o mês m . O componente aleatório é selecionado a partir da abordagem de *bootstrapping*. Um trabalho similar na geração de séries sintéticas é proposto em [14].

A técnica de *bootstrapping* consiste na seleção aleatória, com reposição, de um elemento da série de resíduos associada ao m -ésimo modelo, $m = 1, \dots, 12$. A seleção do componente estocástico é realizada utilizando um gerador de números pseudo-aleatórios com distribuição uniforme.

Neste artigo, para a inicialização da geração da série sintética foram consideradas as médias mensais do histórico de vazões. São geradas séries sintéticas equivalentes a 2000 anos, sendo os primeiros 100 anos desprezados para eliminar o efeito da inicialização.

4 RESULTADOS EMPÍRICOS

Esta seção apresenta os resultados para a geração de series sintéticas de vazões médias mensais como base nos modelos de inteligência computacional FIS e MLP. Para validar a proposta, neste artigo foram consideradas as séries de vazões das usinas de Furnas e Peixoto, pertencentes à cascata do Rio Grande.

As séries de vazões compreendem o período de Janeiro de 1931 a Dezembro de 2008. Para ambas séries, os históricos foram divididos em dois subconjuntos. O primeiro, composto pelas observações de Janeiro de 1931 a Dezembro de 1998, foi utilizado para o ajuste dos parâmetros dos modelos. O segundo subconjunto, Janeiro de 1999 a Dezembro de 2008, foi usado para fins de validação/identificação dos modelos ajustados. Ou seja, os últimos 10 anos foram utilizados para a obtenção dos critério BIC e MSE, usados para a seleção dos modelos FIS e MLP, respectivamente. A Tabela 1 apresenta a estrutura obtida para cada um dos modelos.

Tabela 1: Estrutura dos Modelos FIS e MLP.

Mês	FIS				MLP			
	No. de Entradas		Regras		No. de Entradas		Neurônios	
	Furnas	Peixoto	Furnas	Peixoto	Furnas	Peixoto	Furnas	Peixoto
1	1	1	1	2	3	6	6	8
2	1	1	3	2	4	5	7	7
3	1	1	2	2	3	3	5	6
4	1	1	2	2	5	4	9	7
5	1	1	3	2	3	4	7	6
6	2	2	2	2	2	3	5	6
7	2	1	3	2	3	2	6	5
8	1	1	1	2	5	5	9	8
9	1	1	1	2	6	6	10	11
10	1	1	1	2	4	6	7	9
11	1	3	1	2	2	3	5	5
12	1	1	1	1	5	4	10	8

A Figura 2 mostra as distribuições de frequência relativa das séries históricas e geradas pelos modelos FIS e MLP, para as usinas de Furnas e Peixoto. Nota-se que, para as duas usinas, as vazões com valores altos ocorrem com menor frequência e, consequentemente, são muito menos prováveis do que as vazões baixas. Pode-se observar que, a distribuição de frequência das séries geradas pelos modelos FIS e MLP foram capazes de reproduzir de forma adequada o comportamento histórico das vazões dos postos de Furnas e Peixoto.

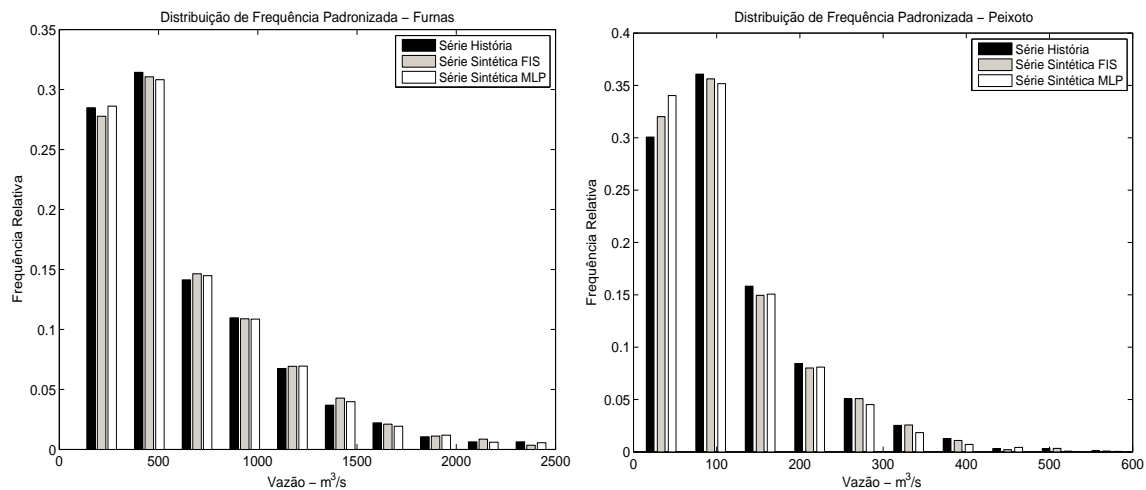


Figura 2: Distribuição de Frequências Relativas - Usinas de Furnas e Peixoto.

As Figuras 3 e 4 mostram as estatísticas descritivas mensais para as usinas de Furnas e Peixoto, respectivamente. As médias mensais tanto para Furnas como para Peixoto, ambos modelos apresentaram valores próximos dos valores das médias históricas de vazões. Já no caso do desvio-padrão, para o posto de Furnas, os valores obtidos pelo modelo MLP estão mais próximos dos valores da série histórica. No caso de Peixoto, os modelos FIS e MLP forneceram valores similares. Analisando as medidas de assimetria, de um modo geral, para Furnas, o modelo FIS apresentou valores menores que os coeficientes de assimetria histórica. A rede MLP obteve valores mais próximos do histórico. No caso da série histórica de Peixoto, o modelo FIS ajustou-se melhor aos coeficientes de assimetria histórico. Nota-se que para o posto de Peixoto há um aumento na assimetria nos meses de Outubro e Novembro que retrata o fim do período de seco e início do período úmido. Finalmente, a medida de curtose mostrou resultados similares para ambos modelos em relação ao coeficiente de curtose obtida da série histórica.

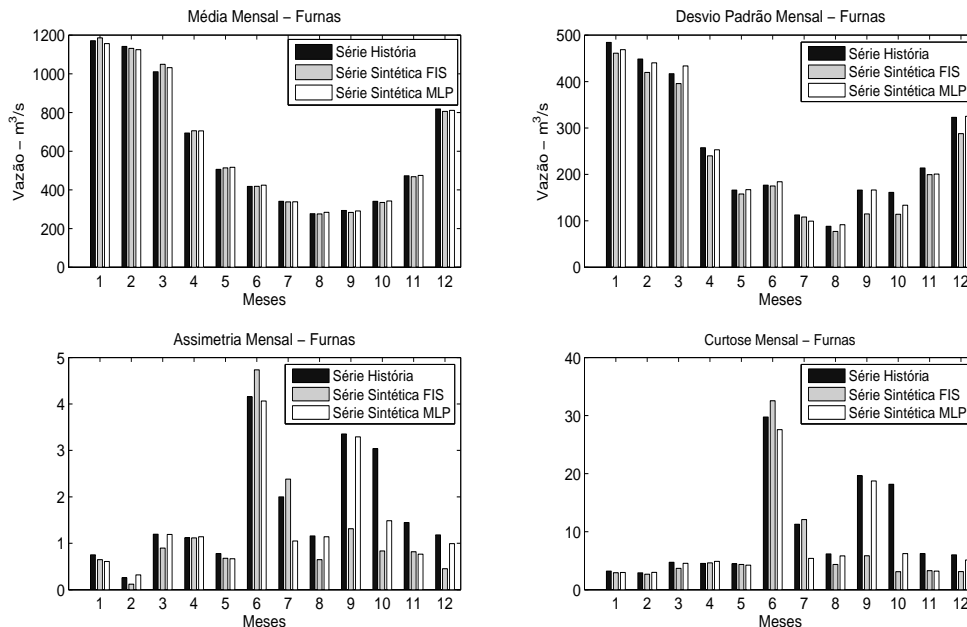


Figura 3: Estatísticas Descritivas - Furnas.

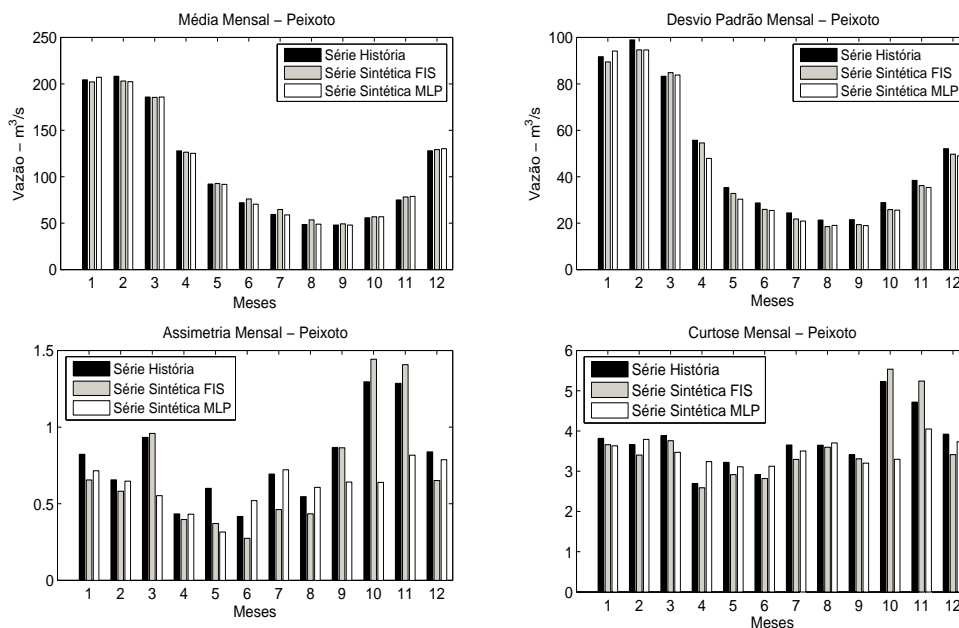


Figura 4: Estatísticas Descritivas - Peixoto.

As Figuras 5 e 6 apresentam os gráficos BoxPlot com base mensal para os postos em análise neste trabalho. Observando estes resultados é possível notar o desempenho favorável para o modelo MLP no caso do posto de Furnas. Por outro lado o modelo FIS mostrou-se extremamente competitivo no posto de Peixoto. Neste gráfico BoxPlot foi possível verificar o comportamento do gerador de séries sintéticas em cada mês do ano sendo possível averiguar que em meses, como por exemplo, março o modelo FIS estimou vazões acima da média levando em consideração o posto de Furnas. Uma análise similar pode ser realizada para o modelo de rede neural no caso da usina de Peixoto.

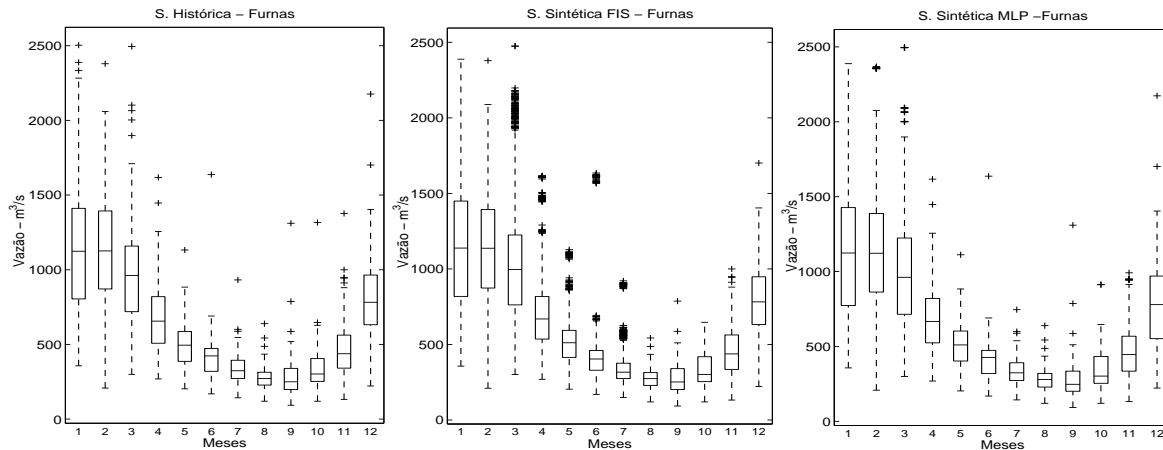


Figura 5: BoxPlot Mensal - Furnas.

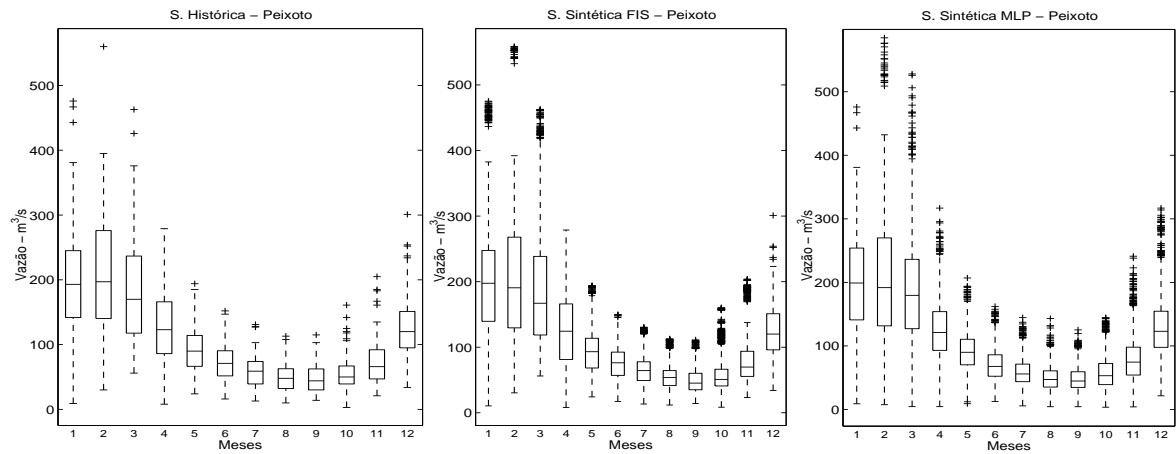


Figura 6: BoxPlot Mensal - Peixoto.

Em geral, os dados sintéticos resultantes da aplicação dos modelos FIS e MLP foram capazes de replicar o histograma observado, preservar a estrutura de autocorrelação, bem como replicar as características gerais da estatística de séries temporais. No entanto, como verificado na Figura 7, o qq-plots mostram que a principal dificuldade de ambos os modelos é a replicação dos maiores valores de vazões (períodos de cheia) observados durante os anos oitenta de dados históricos.

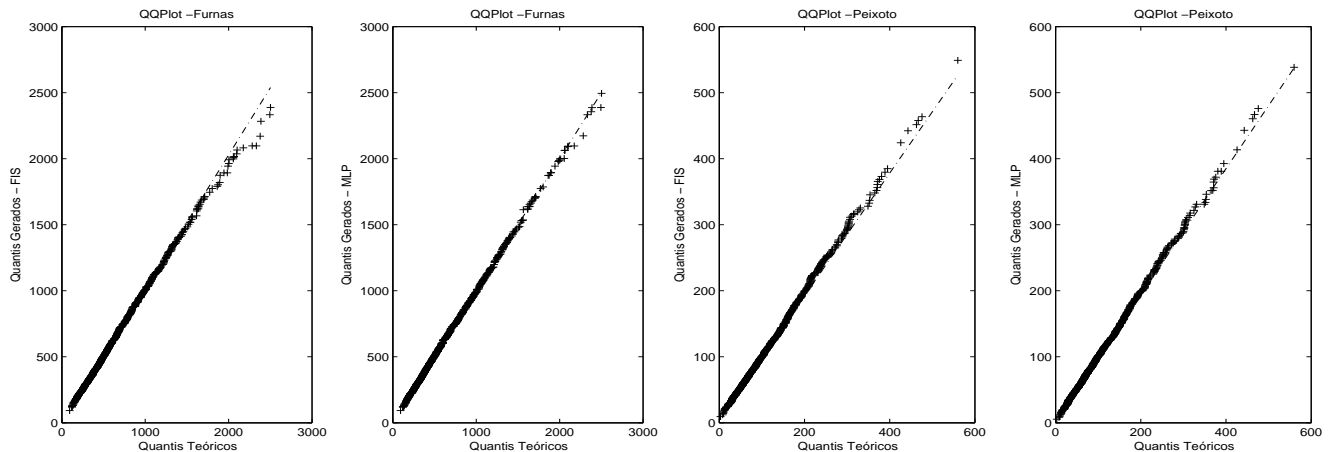


Figura 7: QQPlot Mensal - Furnas/Peixoto.

De forma geral, os modelos de geração de séries sintéticas considerados neste trabalho apresentaram resultados capazes de reproduzir de forma adequada o comportamento estatístico das vazões das usinas hidroelétricas tratadas neste estudo de caso.

5 CONCLUSÕES

Os resultados preliminares apresentados neste artigo mostram que os modelos baseados em regras fuzzy e as redes neurais são potenciais abordagens para a geração de séries sintéticas.

Em geral, as estatísticas descritivas para todos os meses foram devidamente replicadas. Por outro lado, os modelos encontraram dificuldades em replicar os coeficientes de assimetria e curtose para alguns meses das séries de vazões. Deve ser ressaltado que os modelos FIS e MLP não consideram a hipótese de normalidade dos dados, como é feito pela maioria dos modelos de séries temporais. Os resultados forneceram evidências que apontam que os modelos utilizados são uma alternativa adequada para a geração de séries sintéticas de vazões mensais e que reproduzem as principais características estatísticas das séries observadas.

Como proposta futura, os autores pretendem desenvolver estudos comparativos com outros modelos da literatura, bem como o desenvolvimento de testes estatísticos para a validação das séries sintéticas e a análise de outras características hidrológicas, incluindo a distribuição anual e correlação pluri-anual.

6 AGRADECIMENTOS

Os autores agradecem o suporte financeiro das Empresas do Setor Elétrico Brasileiro envolvidas no projeto P&D intitulado ODIN - Modelo de Otimização de Despacho Hidrotérmico, ANEEL PE-0391-0108/2009, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq.

REFERÊNCIAS

- [1] L. Martinez. “Políticas de controle malha fechada e malha aberta no planejamento da operação energética de sistemas hidrotérmicos”. Ph.D. thesis, Faculdade de Engenharia Elétrica e de Computação - Universidade Estadual de Campinas, 2001.
- [2] R. Ballini. “Análise e Previsão de Vazões Utilizando Modelos de Séries Temporais, Redes Neurais e Redes Neurais Nebulosas”. Ph.D. thesis, Faculdade de Engenharia Elétrica e de Computação - Universidade Estadual de Campinas, 2000.
- [3] D. Wilks. “Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain”. *Agricultural and Forest Meteorology*, vol. 96, pp. 85–101, 1999.
- [4] M. Mondal and S. Wasimi. “Generating and forecasting monthly flows of the Ganges river with PAR model”. *Journal of Hydrology*, vol. 323, no. 1-4, pp. 41–56, May 2006.
- [5] D. Lettenmaier. “Synthetic streamflow forecast generation”. *Hydraul. Eng. vASCE*, vol. 110, pp. 277–289, 1984.
- [6] J. D. Salas, J. W. Delleur, V. Yevjevich and W. L. Lane. “Applied modelling of hydrologic time series”. *Water Resources Publications*, p. 484, 1980.
- [7] T. Andersen, H. E. Chung and B. Sorensen. “Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study”. *Journal of Econometrics*, vol. 91, pp. 61–87, 1999.
- [8] V. Yevjevich. “Generation of Hydrologic Samples: Case Study of the Great Lakes-Hydrology Paper 72”. Technical report, Colorado State University, Fort Collins, CO., 1975.
- [9] Y. Jia and T. B. Cluver. “Bootstrapped artificial neural networks for synthetic flow generation with a small data sample”. *Journal of Hydrology*, vol. 331, pp. 580–590, 2006.
- [10] S. Chiu. “A cluster estimation method with extension to fuzzy model identification”. In *Proceedings of The IEEE International Conference on Fuzzy Systems*, volume 2, pp. 1240–1245, June 1994.
- [11] T. Takagi and M. Sugeno. “Fuzzy Identification of Systems and Its Applications to Modeling and Control”. *IEEE Transactions on Systems, Man and Cybernetics*, , no. 1, pp. 116–132, January/February 1985.
- [12] G. Schwarz. “Estimating the dimension of a model”. *Annals Of Statistics*, vol. 6, no. 2, pp. 461–468, 1978.
- [13] P. A. Morettin and C. M. C. Toloí. *Análise de Séries Temporais*. Edgard Blücher Ltda, São Paulo, SP, first edition, 2004.
- [14] R. G.-B. J.C. Ochoa-Rivera and J. Andreu. “Multivariate synthetic streamflow generation using a hybrid model based on artificial neural networks”. *Hydrology and Earth System Science*, vol. 6, no. 4, 2002.