Annual Journal of Hydraulic Engineering, JSCE, Vol.51, 2007, February

SYNTHETIC GENERATION OF MONTHLY RESERVOIR INFLOWS BY AN INPUT DELAYED NEURAL NETWORK

Camilo A. S. de FARIAS[1], Koichi SUZUKI[2], Akihiro KADOTA[3] and Alcigeimes B. CELESTE[4]

[1]Member of JSCE, Ph.D. Student, Dept. of Civil and Environmental Engineering, Ehime University
(3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan)
E-mail: camiloallyson@yahoo.com.br
[2]Fellow of JSCE, Dr. of Eng., Professor, Dept. of Civil and Environmental Engineering, Ehime University
(3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan)
[3]Member of JSCE, Dr. of Eng., Associate Professor, Dept. of Civil and Environmental Engineering, Ehime University
(3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan)
[4]Post-Doctoral Fellow, Dept. of Civil Engineering, Federal University of Campina Grande
(Av. Aprígio Veloso, 882, Campina Grande 58.109-970, Paraíba, Brazil)

An input delayed neural network (IDNN) for synthetic inflow generation is presented to establish monthly inflow scenarios for the reservoir that supplies water to the city of Matsuyama, Japan. IDNNs are dynamic networks capable of accounting for nonlinearities and representing temporal information of input sequences. In this study, the IDNN model relates the two previous reservoir inflows in order to estimate the current inflow. The inflow scenarios will be used as input to optimization models in order to construct reservoir operation policies. Twenty years of historical inflows were used for calibrating the IDNN and a new 20-year synthetic series was generated. Besides the comparison with the IDNN-generated inflows, the statistics of historical series were also compared with those of synthetic series generated by a second-order autoregressive (AR-2) model. The IDNN model proved to be capable of preserving the main statistical characteristics of the historical series.

Key Words **:** *Input delayed neural networks, inflow generation, synthetic scenarios.*

1. INTRODUCTION

The scarcity of water is a world-wide problem and it has been aggravated mainly in regions with increasing population. Such areas have limited economic development and a great need for immediate solutions. As a consequence, many researches have been carried out seeking for more efficient management of the water resources. One such research has been conducted in the city of Matsuyama, Japan. This city has periodical problems with the lack of water, an issue that can be tackled by reservoir operation optimization models. For the implementation of these models, it is frequently necessary to use future reservoir inflow scenarios[1]. The generation of inflow scenarios is generally carried out by models that try to produce synthetic data having statistical characteristics as close as possible to the historical inflow series. This study uses a model based on the well-known artificial neural networks, which have been applied successfully to solving reservoir operating problems in Matsuyama[2), 3)]. The network employed is an input delayed artificial neural network (IDNN). IDNNs are dynamic artificial neural networks capable of accounting for nonlinearities and representing temporal information of input sequences, which make them very suitable for time series modeling[4)].

This work investigates the use of an IDNN model for the synthetic generation of monthly inflows to the reservoir that supplies Matsuyama. Besides the comparison with the historical series, the statistics of the IDNN-generated inflows are also compared with those of synthetic series generated by a second-order autoregressive (AR-2) model.

2. INPUT DELAYED NEURAL NETWORK (IDNN) MODEL

IDNNs contain two components: a memory and an associator. The memory is responsible for holding the past information, which in this study is composed of the two previous inflows or inputs. The associator is a multilayer perceptron network that relates the memory with the desired output, i.e., the current inflow. Thus, the memory component represents the temporal information and the associator accounts for the nonlinearity, making the IDNNs very suitable for time series modeling.

(1) Architecture

The architecture of the network for each month is formed by the input layer, one hidden layer and the output layer. The input layer is composed of two neurons, which are the two previous inflows. The number of neurons in the hidden layer is determined based on a trial-error procedure. The best training results were achieved with five neurons in the hidden layer. The current inflow is the single neuron of the output layer.

(2) Topology

For neural networks, not only the way neurons are implemented but also how their interconnections (topology) are made is important. In this study the network topology is constrained to be feed-forward, i.e., the connections are allowed from the input layer to the hidden layer and from the hidden layer to the output layer. **Figs. 1** and **2** illustrate the network topology of this study and the details of a neuron in the hidden layer, respectively.

In this network, each element of the input vector is connected to each neuron in the hidden layer. The *i*th neuron in the hidden layer has a summation that gathers its weighted inputs and bias to form its own scalar output or induced local field. Each induced local field is submitted to an activation function so that they become the inputs of the output layer. The unique neuron in the output layer also has a summation that gathers its weighted inputs (from the hidden layer) and bias to form its induced local field. This induced local field is then submitted to the neuron activation function and becomes the final output or current inflow.

(3) Activation functions

Continuous and differential functions are necessary for relating inputs and outputs of the artificial neural networks. According to Haykin[5] the sigmoid function is a good activation function for each neuron due to its generally accepted behavior. The tan-sigmoid function is chosen as the activation function for the hidden neurons. For the output layer neuron, a linear activation function is used.

(4) Training process

The original data (input and desired outputs) are conveniently standardized and then scaled before the training in order to improve the efficiency of the IDNN[6]. The standardization process consists of removing seasonality in the mean and variance. The scaling function scales the inputs and targets of the IDNN so that they fall in the range [-1,1].

The training is performed by a back-propagation algorithm which has been successfully applied to water resources systems. In this approach, the Scaled Conjugate Gradient (SCG) method is used for the back-propagation. A detailed explanation of the SCG method is provided by Moller[7]. The network training is supervised, i.e., the series of weights between the neurons and the bias are adjusted through the iterations (epochs) in order to fit the series of inputs to another series of known outputs. The training also

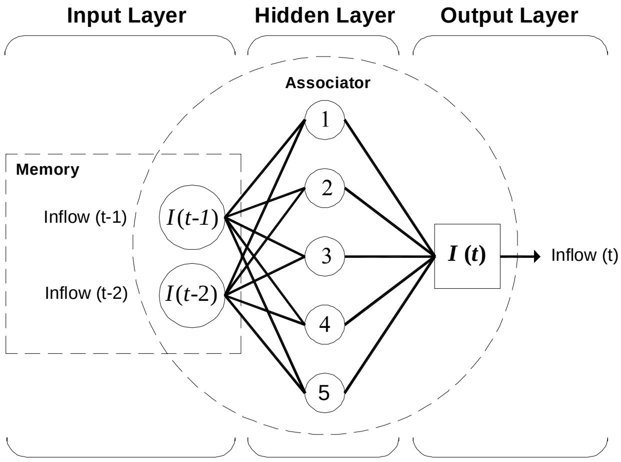
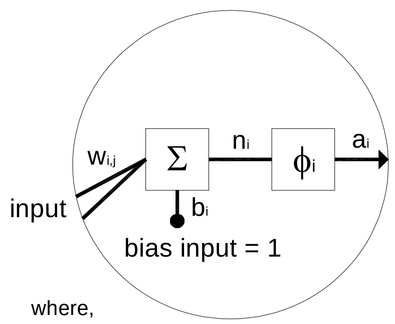


Fig. 1 Topology of the IDNN.



where,

$w_{i,j}$ = weight (*i* indicates the destination neuron of the weight, and *j* indicates which is the input for that weight);
b_i = bias of the *i*th neuron;
n_i = induced local field of the *i*th neuron;
ϕ_i = activation function of the *i*th neuron;
a_i = output of the *i*th neuron;

Fig. 2 Details of a neuron in the hidden layer.

occurs in the batch mode. In this mode the weights and biases are updated only after the entire training set has been applied to the network.

The convergence of the training process occurs when the mean squared error between IDNN outputs and desired results is less than a specific minimum value.

(5) Generation of the synthetic series

The synthetic monthly inflows are obtained by two components: one is deterministic and the other is stochastic.

a) Deterministic component

The IDNN results are the deterministic component for the generation of synthetic inflows. The IDNN series of outputs for each month of the year are obtained from the calibrated IDNN model operation as follows:

$$\mathrm{a}_i(t) = \left[\sum_{j=1}^{IN=2} x_j(t) \times w_{i,j}\right] + b_i \tag{1}$$

$$o(t) = \phi_2\left\{\left\{\sum_{i=1}^{HN=5} \phi_1\left[\mathrm{a}_i(t)\right] \times w_{h,i}\right\} + b_h\right\} \tag{2}$$

$$O = \{o(1), o(2), \cdots, o(n)\} \tag{3}$$

where t is the time index; i is the neuron index in the hidden layer; $\mathrm{a}_i(t)$ is the induced local field of the ith neuron in the hidden layer at time t; $x_j(t)$ is the value of the jth neuron in the input layer at time t; IN is the number of input neurons; $w_{i,j}$ is the weight between the jth input neuron and the destination neuron i; b_i is the bias of the ith neuron in the hidden layer; $o(t)$ is the IDNN output at time t; ϕ_2 is the linear activation function for the neuron in the output layer; HN is the number of hidden neurons; ϕ_1 is the sigmoid activation function for all neurons in the hidden layer; $w_{h,i}$ is the weight between the ith input neuron (from hidden layer) and the destination neuron h; b_h is the bias of the output layer neuron; O is the set of results from the IDNN model; and n is the number of months to be generated by the model. The deterministic component is then obtained by applying the inverse of the scaling function:

$$o'(t) = \left(\frac{o(t)+1}{2}\right) \times \left(k_{\max} - k_{\min}\right) + k_{\min} \tag{4}$$

$$O' = \{o'(1), o'(2), \cdots, o'(n)\} \tag{5}$$

in which $o'(t)$ is the output for the inverse of the scaling function at time t; $k_{\max}$ and $k_{\min}$ are the maximum and minimum values in the series of known outputs used for calibrating the IDNN, respectively; and O' is the deterministic component.

b) Stochastic component

When the set of standardized IDNN outputs and observed results are compared, the difference between them leads to a residuals series. Statistical analysis of the residuals shows that they can be modeled as gamma distributions. A fitting of the residuals series is provided, and the posterior random generation of these values is the stochastic component of the model.

c) Inflow series generation

The synthetic monthly inflow series is found by the following equations:

$$r = \{r(1), r(2), \cdots, r(n)\} \tag{6}$$

$$Inf(t) = \left[o'(t) + r(t)\right] \times S(m) + M(m) \tag{7}$$

$$INF = \{Inf(1), Inf(2), \cdots, Inf(n)\} \tag{8}$$

where r is the series of random generated residuals obtained from a properly fitted gamma distribution; $r(t)$ is the residual at time t; m is the index for months of the year (January, February, …, December); $S(m)$ is the standard deviation for month m; $M(m)$ is the mean for month m; $Inf(t)$ is the synthetic monthly inflow at time t (this equation contains the inverse functions of the standardization process); and INF is the series of synthetic monthly inflows.

Like the fitting and generation of residuals series, the IDNN model was implemented in Matlab environment.

3. GENERATION OF THE SYNTHETIC SERIES

The IDNN model was applied for generating synthetic monthly inflows to the Ishitegawa Dam reservoir. This reservoir supplies water to the city of Matsuyama, located in Ehime, Japan. The reservoir is also used for irrigation and flood control. Twenty years of historical monthly inflows were used for calibrating the IDNNs for each month of the year. After the determination of the deterministic component by **Eqs. (1-5)**, the synthetic series were obtained by the random generation of the fitted gamma distribution applied to **Eqs. (6-8)**.

In order to check the performance of the IDNN model, 100 synthetic series of monthly inflows (each 20-year long) were generated and its statistics were compared with those from the historical data. In addition, a second order autoregressive model (AR-2) was applied for comparison. Same pre-processing of data series was used for both AR-2 and IDNN model.

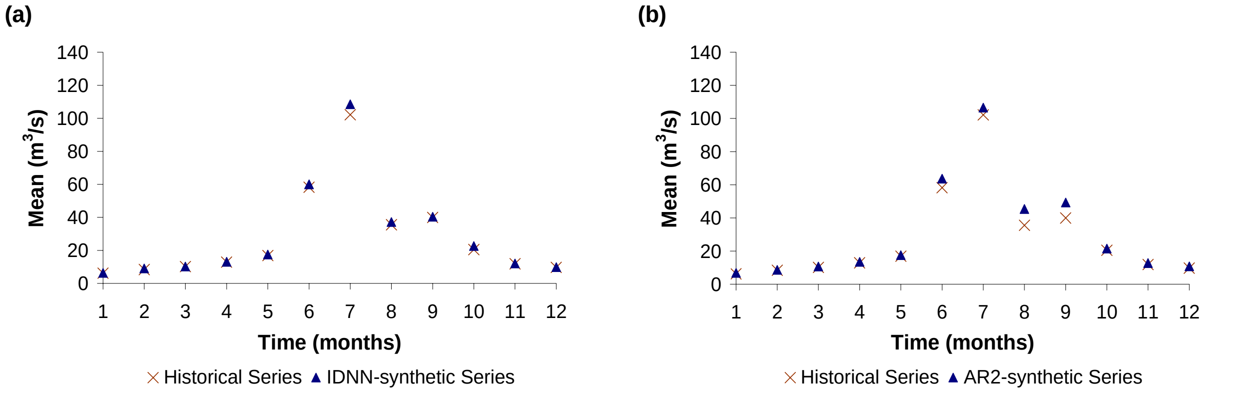


Fig. 3 Comparison of monthly mean of the historical series with (a) IDNN and (b) AR-2 synthetic series.

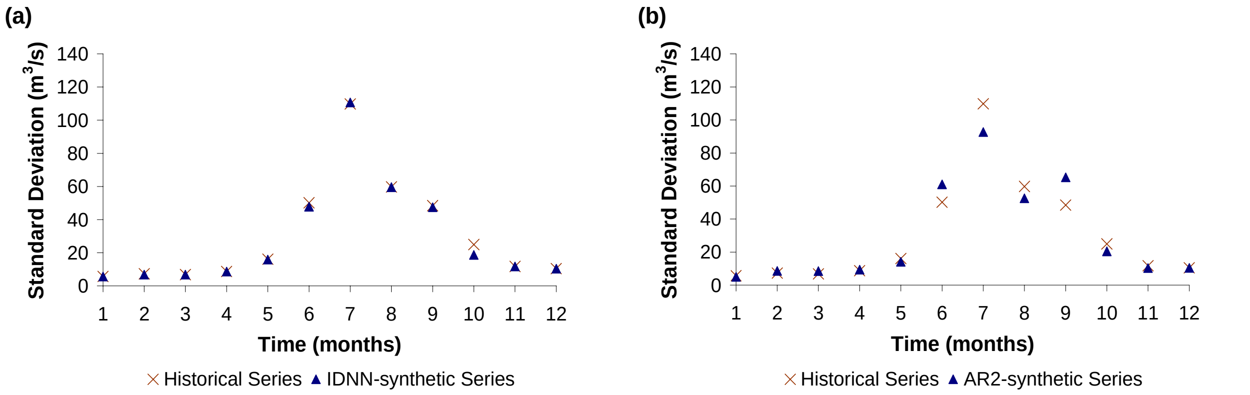


Fig. 4 Comparison of monthly standard deviation of the historical series with (a) IDNN and (b) AR-2 synthetic series.

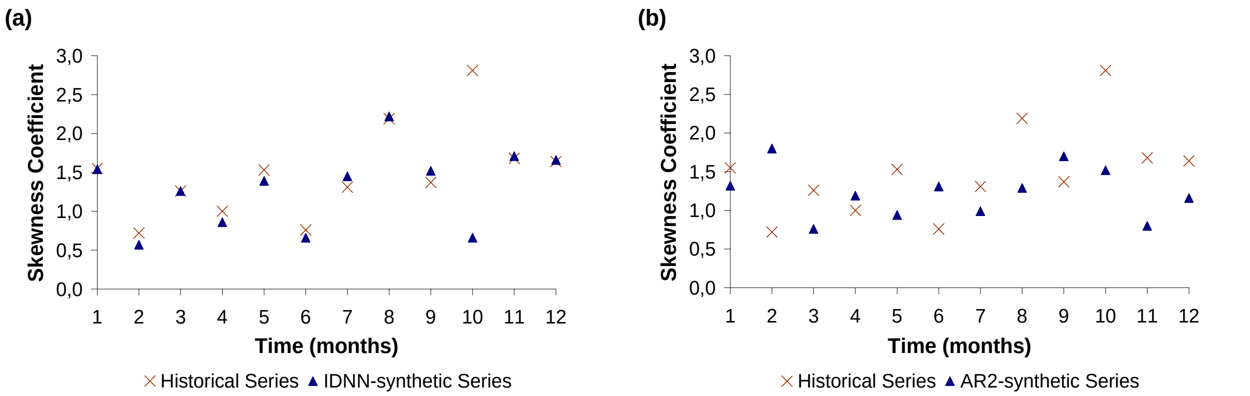


Fig. 5 Comparison of monthly skewness coefficient of the historical series with (a) IDNN and (b) AR-2 synthetic series.

Statistical characteristics such as mean, standard deviation and skewness coefficient of the IDNN and AR-2 synthetic data were compared with those from the historical series. **Figs. 3-5** illustrate graphs comparing monthly mean, standard deviation and skewness coefficient of the historical series with the ones found by the IDNN and AR-2 synthetic series. Annual statistics for historical and synthetic series are presented in **Table 1**. The histograms of the

Table 1 Annual statistics for historical and synthetic series.

Series	Mean (m^3/s)	Standard Deviation (m^3/s)	Skewness Coefficient
Historical	332.52	222.63	1.77
IDNN	346.82	215.63	1.99
AR-2	367.07	171.23	0.95

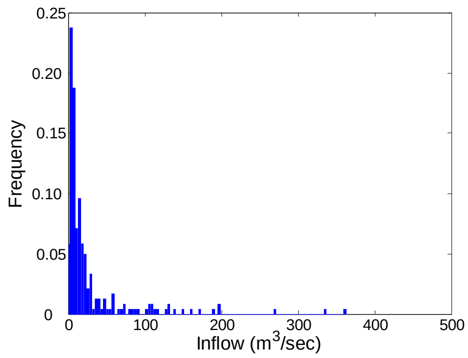


Fig. 6 Histogram of the historical series.

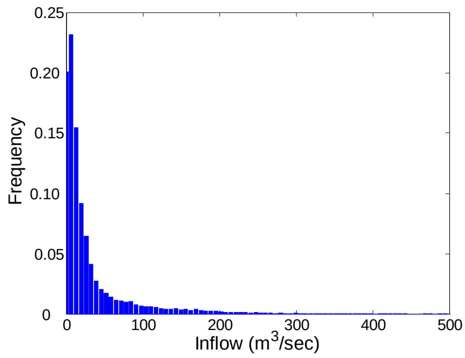


Fig. 8 Histogram of the AR-2 synthetic series.

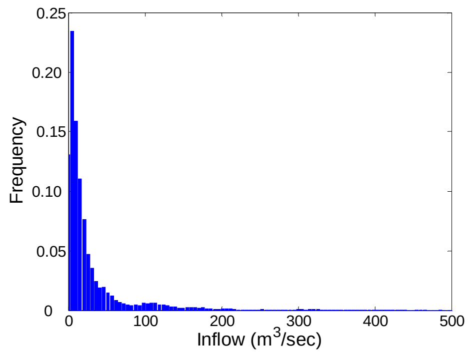


Fig. 7 Histogram of the IDNN synthetic series.

historical and synthetic series are presented in **Figs. 6-8**.

The results showed that the monthly and annual statistics found by the synthetic-IDNN series were very close to those showed in the historical series.

The AR-2 model also generated synthetic series with satisfactory performance for most monthly statistics. However, poor accuracy was found in the monthly skewness coefficients and annual statistics. Both models showed histograms similar to the one obtained by the historical series.

Comparing all statistics from the IDNN-generated series with the ones from the AR-2 model, it can be noticed that the IDNNs' capabilities of identifying the nonlinear trends among the hydrologic variables and representing temporal information overcomes the pure autoregressive technique. Since the IDNN procedure was capable of fitting the historical series better than the AR-2 model, it can be said that the proposed approach may produce more reliable synthetic series and hence more consistent data for testifying reservoir optimization techniques and solving hydrological problems.

4. GENERATION OF SYNTHETIC INFLOW SCENARIOS

Optimization techniques for the operation of Ishitegawa Dam are going to be tested with possible reservoir inflow scenarios.

Synthetic inflow scenarios were generated by the IDNN (described in Section 2) and AR-2 models to show the different possibilities of inflow occurrence. As an example, **Figs. 9-10** show data for a 5-scenario

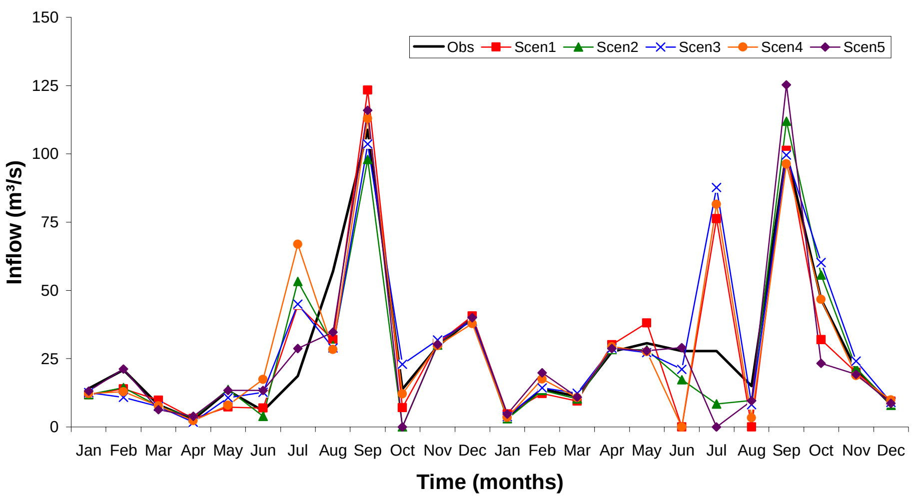


Fig. 9 Generated scenarios from 1989 until 1990 by the IDNN model.

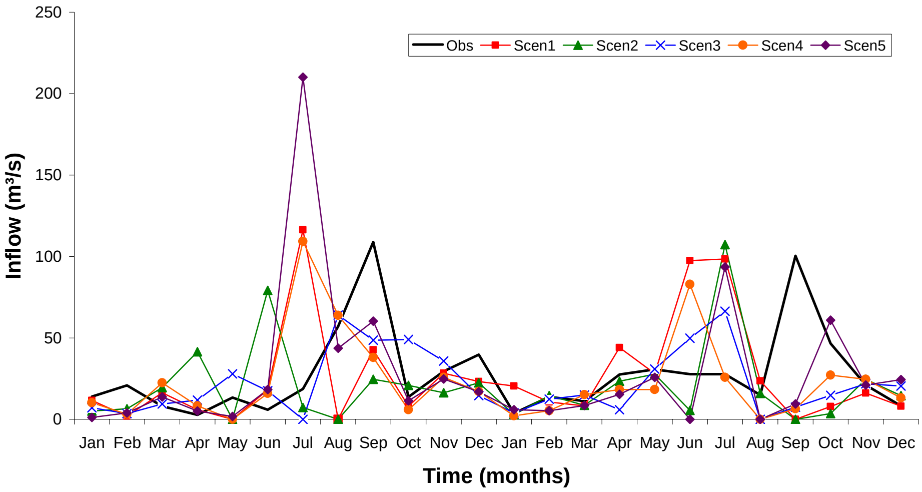


Fig. 10 Generated scenarios from 1989 until 1990 by the AR-2 model.

generation for the years of 1989-1990 by IDNN and AR-2 models, respectively. The observed inflow is also provided for comparison.

Any number of possible synthetic scenarios can be carried out by the IDNN or AR-2 models. However, the annual and monthly statistics (Section 3) obtained from the scenarios generated by both IDNN and AR-2 models suggest that the first may produce more trustworthy inputs to the optimization models than the AR-2.

5. CONCLUSIONS

An input delayed artificial neural network was applied for generating synthetic monthly inflow scenarios for Ishitegawa Dam, which is the reservoir that supplies the city of Matsuyama, in Ehime Prefecture. The scenarios will be used by optimization techniques that have been developed for the reservoir operation.

The monthly and annual statistics (mean, standard deviation and skewness coefficient) obtained by the IDNN model were very close to the ones presented in the historical data. Moreover, the synthetic inflow histogram was similar to that from the observed data.

The comparison between IDNN and AR-2 models indicates that the IDNNs' capabilities of accounting for nonlinearities and representing temporal information produced more reliable synthetic series than the statistical method. As conclusion, the results suggest that the IDNN model is suitable for generating the synthetic monthly inflow scenarios needed by the optimization techniques for the operation of Ishitegawa Dam.

REFERENCES

1) Celeste, A. B., Suzuki, K., Kadota, A., Farias, C. A. S.: Stochastic generation of inflow scenarios to be used by optimal reservoir operation models, *Annual Journal of Hydraulic Engineering*, JSCE, Vol. 48, pp. 451-456, 2004.
2) Farias, C. A. S., Celeste, A. B., Sakata, Y., Kadota, A., Suzuki, K.: Use of monte carlo optimization and artificial neural networks for deriving reservoir operating rules, *Annual Journal of Hydraulic Engineering*, JSCE, Vol. 50, pp. 25-30, 2006.
3) Farias, C. A. S., Celeste, A. B., Kadota, A., Suzuki, K.: ANN-against interpolation-based implicit stochastic optimization for deriving reservoir operating rules, *Proceedings of the 15th Congress of APD-IAHR*, ITT Madras, Chennai, India, pp. 633-639, 2006.
4) Coulibaly, P., Anctil, F., Bobée, B.: Multivariate reservoir inflow forecasting using temporal neural networks, *Journal of Hydrologic Engineering*, ASCE, Vol. 6, pp. 367-376, 2001.
5) Haykin, S.: *Neural networks: a comprehensive foundation*, Prentice Hall, Inc., 2.ed, New Jersey, 1999.
6) Salas, J. D.: Analysis and modeling of hydrologic time series, In: Maidment, D. R.: *Handbook of Hydrology*, McGraw-Hill, Inc., New York, pp. 19.1-19.72, 1993.
7) Moller, M. F.: Scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, Vol. 6 (4), pp. 525-533, 1993.

(Received September 30, 2006)