Forecasting of Daily Inflows Based on Regressive Neural Networks

Shin, Hyun-Suk
Assist. Prof., Department of Civil Engineering, Pusan University, Pusan, Korea

Kim, Tae-Woong
Grad. Student, Department of Civil Engineering, University of Arizona, Tucson, AZ, USA

Kim, Joong-Hoon
Assoc. Prof., Department of Civil and Environmental Engineering, Korea University, Seoul, Korea

ABSTRACT: The daily inflow is apparently one of nonlinear and complicated phenomena. The nonlinearity and complexity make it difficult to model the prediction of daily flow, but attractive to try the neural networks approach which contains inherently nonlinear schemes. The study focuses on developing the forecasting models of daily inflows to a large dam site using neural networks. In order to reduce the error caused by high or low outliers, the back propagation algorithm which is one of neural network structures is modified by combining a regression algorithm. The study indicates that continuous forecasting of a reservoir inflow in real time is possible through the use of modified neural network models. The positive effect of the modification using the regression scheme in BP algorithm is showed in the low and high ends of inflows.

1 INTRODUCTION

In order to operate reservoirs efficiently for both water utilization and flood control purposes, it is necessary to predict the inflow to the reservoirs in a precise manner. The daily inflow has an apparent nonlinear and complicated physical structure which makes difficult to model it, but daily runoff forecasting is used for reference data in planning of supply and active storage of reservoirs.

Neural networks (NNs) is a computational framework consisting of massively connected simple processing units which called neurons. The following advantages which can be achieved by using neural networks for rainfall-runoff relationship analysis have been showed by various researches recently; 1) Application of NNs does not require a prior knowledge of the process because NNs has black-box properties; 2) NNs easily converges to the optimal solution and does not need any assumptions; 3) NNs has inherently the property of nonlinearity; 4) NNs can have the multiple inputs having the different characteristics, this can make NNs have the time-space variant property; 5) NNs has the adaptability to the change of problem environment, therefore, using a NNs, several cases can be simulated for the same knowledge. The reasons as mentioned above make NNs have been widely studied and applied for solving the water resources problems.

French et al.(1992) developed a neural networks to forecast rainfall intensity fields in both space and time domains. Tang and Fishwick (1993) showed the applicability of neural networks as one of the models for the time series forecasting and compared the Box-Jenkins method against the neural network method for long and short-term memory series. Zhu et al.(1994) proposed a new method to forecast runoff using neural networks and compared with the fuzzy method. Smith and Eli (1995) used spatial distribution rainfall patterns as input data to measure runoff. Thirumalaiah and Deo(1998a and 1998b) forecasted water levels of which the occurrence is influenced by physical processes which are highly complex and uncertain. Kim(2000) forecasted daily streamflow at Jindong station in Nakdong river basin using multiplayer neural networks models, and compared the results with those obtained by a multiple regression model.

Up to now, although many NNs models are developed for rainfall-runoff relationship, many of them are limited to forecasting just short-term hydrologic data considering the flood events. In this paper, we try to make the neural network model to forecast the continuous daily inflow to be used effectively for reservoir operation in real time basis. In addition, we suggest a new modified scheme of neural forecasting model to improve the forecasting accuracy for the extremely high and low flow conditions.

2 NEURAL NETWORKS

A neural network is a computational framework consisting of massively connected simple processing nodes, or neurons. The nodes are classified into input, hidden and output nodes. Input nodes receive data from sources external to the network, hidden nodes send and receive data only from other nodes in the network, and output nodes produce data generated by the network which goes out of the system. They are typically interconnected among themselves by weights. The feed forward neural networks provide a general framework for representing nonlinear functional mappings between a set of input variables and a set of output variables. The typical three-layered networks shown in Fig. 1 are based on a linear combination of the input variables which are transformed by a nonlinear activation function.

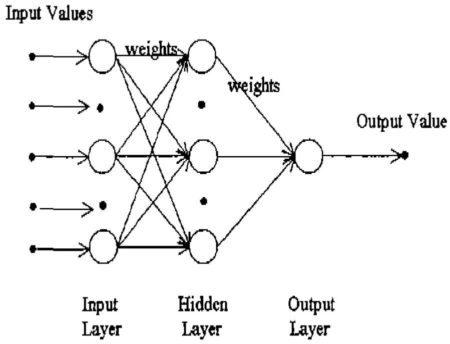


Fig. 1. Scheme of Neural Network Model

In this network there are d inputs, M hidden nodes and C output nodes. The complete explicit expression for the function is obtained by

$$y_k = \tilde{f}\left[\sum_{j=0}^{M} w_{kj} f\left(\sum_{i=0}^{d} w_{ji} x_i\right)\right] \tag{1}$$

where, w_{ji} is a weight in the hidden layer, $f(\cdot)$ is the activation function of the hidden neurons, w_{kj} is a weight in the output layer, and $\tilde{f}(\cdot)$ is the activation function of the output neurons.

Among many NNs' paradigms, the back-propagation algorithm is by far most popular. This algorithm minimizes the sum-squared error of the network by using the steepest descent approach. The NNs' weights and bias are adjusted moving a small step in the direction of the negative gradient of the error function during iteration. This algorithm is based on the error-correction learning rule. Basically, the error back-propagation process consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, an input vector is applied to the nodes of the network, and its effect propagates through the network, layer by layer. Finally, a set of output is produced as the actual response of the network. During the forward pass the synaptic weights of the network are all fixed. The common methods to improve the training and testing accuracy are to use a momentum and an adaptive learning rate in back propagation routine. The reader can refer to Haykin (1994) for more detail description of feed-forward neural networks with back propagation and its improvements.

3 MODIFICATION OF NEURAL NETWORKS

Due to the high variability of daily inflow data containing extremely high values in flood seasons and low values in dry seasons, the neural network model for inflow forecasting usually has made worse accuracy of forecasting due to over-trained and under-trained problems. In order to overcome the problems, this study suggests a modified NN scheme using a regression technique as follows and as shown in Fig. 2. In the modified NN model, first the model is trained using back propagation process and then the necessary weights are adjusted.

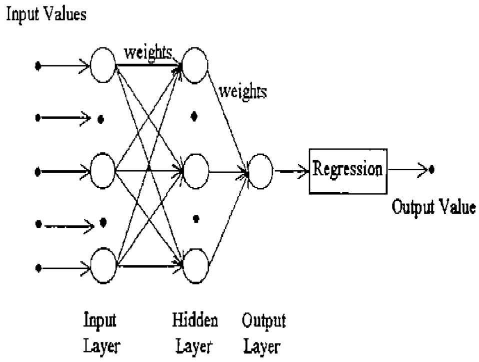


Fig. 2. Schemes of Modified Neural Networks

Then, simple regression parameters are obtained between the trained (estimated) output and corresponding actual observed data. During this, the whole range of data is divided into adequate intervals and the regression parameters are determined for each interval, respectively. Then the forecasting models are modified including the regression parameters directly as shown in Fig 2. In this study, the three cases of rainfall–runoff process are considered as follows, which may be possible in the real situation of reservoir operation; (1) the case available to get point rainfall data from several gages and a inflow into reservoir, (2) the case that mean rainfall over the proposed basin and the inflow can be obtained, and (3) the case with only the inflow data into the reservoir. Then the corresponding forecasting models of daily inflow are expressed as follows:

■ Nonlinear Spatial Rainfall-Runoff Model : R-NSRRM

$$\hat{Q}(T) = \alpha \times fn\begin{bmatrix} R^k(t-1), & R^k(t-2), & \cdots, & R^k(t-n_{Rk}); \\ Q(t-1), & Q(t-2), & \cdots, & Q(t-n_Q) \end{bmatrix} \tag{2}$$

■ Nonlinear Average Rainfall-Runoff Model : R-NARRM

$$\hat{Q}(T) = \beta \times fn\begin{bmatrix} \bar{R}(t-1), & \bar{R}(t-2), & \cdots, & \bar{R}(t-n_R); \\ Q(t-1), & Q(t-2), & \cdots, & Q(t-n_Q) \end{bmatrix} \tag{3}$$

■ Nonlinear Auto-regressive Runoff Model : R-NARM

$$\hat{Q}(T) = \gamma \times fn\left[Q(t-1), \quad Q(t-2), \quad \cdots, \quad Q(t-n_Q)\right] \tag{4}$$

where, $\hat{Q}(t)$ is the forecasted inflow, T is the forecast time ($t + leadtime$), α, β, and γ are linear regression factors, $R^k(t-n_{Rk})$ is the rainfall with n_{Rk} lag time in k site, $\overline{R}(t-n_R)$ is the mean rainfall with n_R lag time, and $fn(\cdot)$ represents the relationship between input and output expressed by Eq. (1) in NNs. As mentioned before, the regression factors are obtained from the training output. Here, the NSRRM is the model using point rainfall data as input in order to consider the spatial distribution of rainfall. NARRM is the model using mean rainfall data so it does not consider the spatial distribution of rainfall. NARM is the model using inflow data only as in ARMAX. The notation R- in Eq. (2)~(4) indicates that the model is modified with regression. Each model is trained and applied for normal, flood, and dry flow seasons, respectively, in order to consider the seasonal variability of daily flow in modeling.

4 STUDY AREA

The project region was the basin of Daechung Dam located in Geum River basin in Korea as shown in Fig. 3. That Daechung Multipurpose Dam is located 150 km upstream from Geum River mouth with Yellow Sea and has a storage capacity of $1,490\times10^6$ m^3 and a drainage area of 4,134 km^2. The daily inflow data for the period from January 1983 to December 1995 are obtained and used in this study.

The temporal variation of data is quite big due to the seasonal variability of meteorological and morphologic characteristics, which needs to be considered. Therefore, the modified NN model have been trained with the input patterns for three distinctive seasons, namely normal(from March to June), flood(from July to September), and dry(from October to Next February) seasons, respectively.

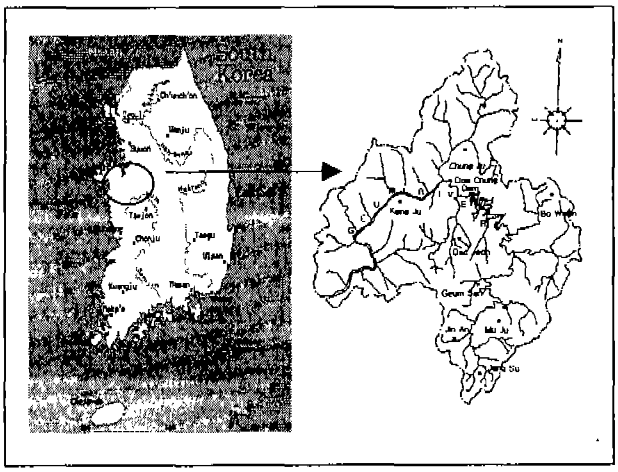


Fig. 3. Geum River Basin and Daechung Dam Subbasin

5 NEURAL NETWORKS IMPLEMENTATION

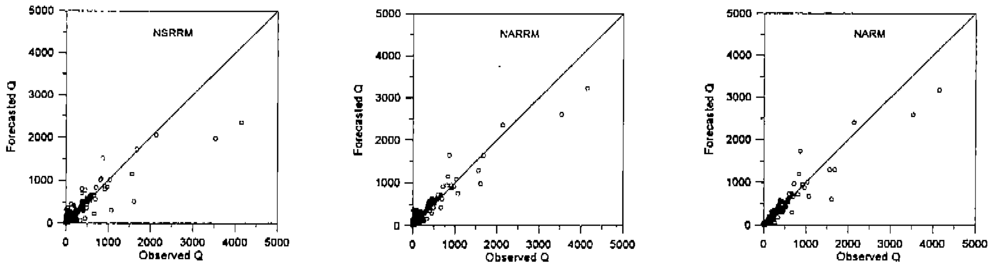
The data form 1983 to 1995 was divided into two parts: one is training pattern including 1983 to 1992 data and the other is testing (forecasting) pattern including 1993 to 1995 data. At first, the data was normalized to the range of [0.05, 0.95] before feeding into the neural network models to improve the efficiency of NNs model empirically. It is a type of preprocessing to prevent the noise and distortion from nonlinear transformation. Then, it was necessary for the output to be denormalized after training and forecasting. The number of hidden neurons was determined experimentally to meet the model accuracy since there has not been any well-defined algorithm for determining the optimal number of hidden neurons.

6 RESULTS OF FORECASTING DAILY INFLOW

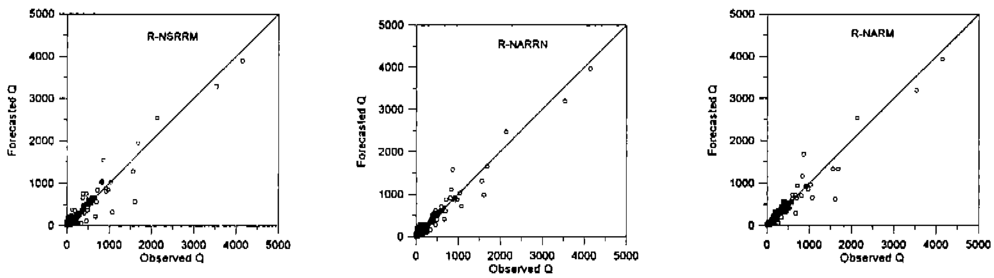
Both the original models NNs (NSRRM, NARRM, and NARM) and the proposed modified models R-NNs (R-NSRRM, R-NARRM, and R-NARM) for forecasting daily inflow are constructed to forecast the 1-Day, 2-Day, 3-Day, and 5-Day ahead inflows. The forecasting efficiency is compared for both models and showed clearly in Fig.4. The result indicates that the R-NNs are superior to the NNs in terms of fitting the low and high inflow data for all cases tested in this study.

Figure 5 compares the validation results of daily inflow forecasting between the NNs and the R-NNs with increasing lead times. In terms of RMSE(root mean square error) and CC(correlation coefficient), the results indicate that the R-NNs result lower RMSE and higher CC than the NNs, which means better forecasting accuracy of the R-NNs. The forecasting accuracy is decreased for all cases as the lead time increases, but the R-NNs still show superiority in the cases with the increased lead time than the NNs do.

In Table 1 and 2, the forecasting results between the NNs and the R-NNs are compared regarding to both peak flow and volumes for year 1993 and 1995, respectively. The ratios of forecasred peak flow to the observed peak flow in percentage are indicated in the blankets. The results indicate that both the peak flow and the inflow volume from the R-NNs are closer to the real values than those from the NNs are. The peak flows estimated by the NNs seem to be underestimated seriously, but are improved by the R-NNS by approximately 30 % in average. According to Table 2, the R-NNs showed closer inflow volumes than the NNs. For the NNs, the inflow volumes are under-estimated in 1993 and over-estimated in 1994 and 1995. The R-NNs considerably adjusts these poorly estimated inflow volumes.

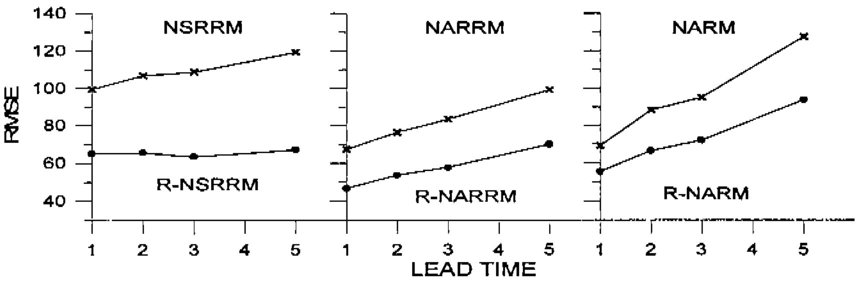


(a) Original NNs Models

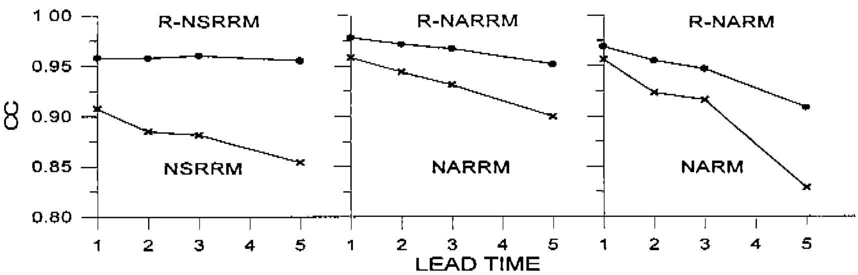


(b) Modified R-NNs Models

Fig. 4. Comparison of Scatter Plot for 1-Day ahead Forecasting Inflow



(a) RMSE



(b) Correlation Coefficient

Fig. 5. Comparison of Forecasting Results Regarding to Lead Time

Among three NNs, the NARRM performed apparently the best. In the R-NNs, R-NARRM shows the best performance in forecasting the reservoir inflow, too. It can be analyzed that the mean rainfall data used in NARRM reduce the noise and deviation contained in the spatial rainfall data since NARRM has a similar trend between its input and output, that is, between hyetograph and hydrograph.

Table 1. Comparison of Peak Inflows from NNs and R-NNs (unit:cms)

Year	Lead Time	NSRRM	R-NSRRM	NARRM	R-NARRM	NARM	R-NARM
1993	1D	1981.8 (56.0%)	3290.9 (93.0%)	2599.0 (73.4%)	3191.2 (90.2%)	2584.9 (73.0%)	3194.8 (90.3%)
	2D	1982.0 (56.0%)	3349.4 (94.6%)	2327.9 (65.8%)	3132.4 (88.5%)	2335.2 (66.0%)	3088.6 (87.3%)
	3D	1893.2 (52.5%)	3225.5 (91.1%)	2189.4 (61.9%)	2992.5 (84.5%)	2079.9 (58.8%)	2969.9 (83.9%)
	5D	1798.5 (50.8%)	2979.6 (84.2%)	1774.6 (50.1%)	2656.2 (75.0%)	1409.3 (39.8%)	2546.3 (71.9%)
1995	1D	2345.3 (56.4%)	3894.4 (93.6%)	3227.9 (77.6%)	3963.4 (95.3%)	3180.2 (76.5%)	3930.6 (94.5%)
	2D	2286.3 (55.0%)	3863.6 (92.9%)	2935.1 (70.6%)	3949.6 (94.9%)	2974.2 (71.5%)	3933.7 (94.6%)
	3D	2326.2 (55.9%)	3963.3 (95.3%)	2979.1 (71.6%)	4067.9 (97.8%)	2767.9 (66.5%)	3952.4 (95.0%)
	5D	2538.9 (61.0%)	4206.2 (101.1%)	2890.3 (69.5%)	4326.2 (104.0%)	2196.9 (52.8%)	3969.2 (95.4%)

(·) : Percentage to the observed values 3539.5cms of 1993 and 4159.5cms of 1995

Table 2. Comparison of Inflow Volume between NNs and R-NNs (unit : 10^9 m^3)

Year	Lead Time	NSRRM	R-NSRRM	NARRM	R-NARRM	NARM	R-NARM
1993	1D	2.8059 (76.1%)	2.9837 (80.9%)	3.2135 (87.2%)	3.8192 (103.6%)	3.3540 (91.0%)	3.5102 (95.2%)
	2D	2.7042 (73.3%)	3.2398 (87.9%)	2.9845 (80.9%)	3.2246 (87.5%)	3.1722 (86.0%)	3.7077 (100.6%)
	3D	2.7549 (74.7%)	3.4298 (93.0%)	2.9799 (80.8%)	3.4499 (93.6%)	3.2897 (89.2%)	3.5638 (96.7%)
	5D	2.6827 (72.8%)	3.6439 (98.8%)	2.8024 (76.0%)	3.2375 (87.8%)	3.1247 (84.7%)	3.4736 (94.2%)
1994	1D	1.1585 (139.8%)	0.7880 (95.1%)	1.3949 (168.3%)	0.8758 (105.7%)	1.1023 (133%)	0.9057 (109.3%)
	2D	1.3089 (157.9%)	0.7277 (87.8%)	0.9772 (117.9%)	0.7198 (86.8%)	1.3669 (164.9%)	0.9584 (115.6%)
	3D	1.5209 (183.5%)	0.7855 (94.8%)	1.1624 (140.2%)	0.7587 (91.5)	1.3884 (167.5%)	0.9765 (117.8%)
	5D	1.8410 (222.1%)	0.8332 (100.5%)	1.3053 (157.5)	0.8041 (97.0%)	1.3372 (161.3%)	0.9465 (114.2%)
1995	1D	1.9556 (120.0%)	1.7695 (108.6%)	2.3410 (143.7%)	1.8739 (115.0%)	2.0240 (124.2%)	1.9003 (116.6%)
	2D	2.2580 (138.6%)	1.7646 (108.3%)	1.8537 (113.7%)	1.7004 (104.3%)	2.2758 (139.6%)	1.9084 (117.1%)
	3D	2.4790 (152.1%)	1.8289 (112.2%)	2.1604 (132.6%)	1.8088 (111.0%)	2.2212 (136.3%)	1.9460 (119.4%)
	5D	2.8244 (173.3%)	1.8563 (113.9%)	2.1283 (130.6%)	1.7750 (108.9%)	2.1525 (132.1%)	1.8992 (116.5%)

(·) : Percentage to the observed values 3.6871× 10^9 m^3 of 1993, 0.8289× 10^9 m^3 of 1994, and 1.6296× 10^9 m^3 of 1995

7 CONCLUSIONS

This paper focused on developing forecasting models for the daily inflows into a large reservoir based on neural networks. Moreover, in an effort to improve the accuracy of high and low flow forecasting, a modified model with regression scheme was suggested and validated comparing with the traditional neural network scheme. Then, the following conclusions could be made.

(1) Three model types were suggested to cover the possible situations which can occur in the real reservoir operation. The models were following; NSRRM : the model which can be used if spatial rainfall data can be obtained for a reservoir basin in real time; NARRM: the model which can be used if spatially averaged rainfall data can be obtained; NARM: the model which can be used with the inflow only without rainfall data.

(2) The modified neural network models, R-NNs, were superior to the traditional neural network models, NNs, for continuous forecasting of reservoir inflows in real time in terms of several measures such as RMSE, CC, peak inflow and inflow volume.

(3) Regarding to the lead times of 1-day, 2-day, 3-day, and 5-day, the R-NNs showed better performance than the NNs did. In addition, the R-NNs showed an acceptable forecasting performance even for 5-day, while the NNs were failed.

(4) The R-NN scheme developed in this study is exected to effectively model the time series with high variability in nature.

8 REFERENCES

French, M.N., Krajewski, W.F., and Cuykendall, R.R. (1992). "Rainfall forecasting in space and time using a neural network." *Journal of Hydrology*, Vol. 137, pp. 1-31.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*, Macmillan College Publishing Co., Ontario, Canada.

Kim, S.W. (2000). "A study on the forecasting of daily streamflow using the multilayer neural networks model." J. of KWRA, Vol. 33, No. 5, pp.537-550.

Lettenmaier, D.P. and Wood, E.F. (1992). "Hydrologic forecasting." *Handbook of Hydrology*, Edited by maidment, D. R., McGraw-Hill, pp. 26.1-26.30.

Smith, J. and Eli, R.N. (1995). "Neural-network models of rainfall-runoff process." *J. of Water Resources Planning and Management*, ASCE, Vol. 121, No. 6, pp. 499-508.

Tang, Z. and Fishwick, P.A. (1993). "Feedforward neural nets as models for time series forecasting." *Journal on Computing*, ORSA, Vol. 5, No. 4, pp. 374-385.

Thirumalaiah, K. and Deo, M.C. (1998a). "River stage forecasting using artificial neural networks." *J. of Hydrologic Engineering*, ASCE, Vol. 3, No. 1, pp. 26-32.

Thirumalaiah, K. and Deo, M.C. (1998b). "Real-time flood forecasting using neural networks." *Computer-Aidded Civil and Infrastructure Engineering*, Blackwell Publishers, Vol. 13, pp. 101-111.

Zhu, M.L., Fujita, M., and Hashimoto, N. (1994). "Application of neural networks to runoff prediction." *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Vol. 3, pp. 205-216.