TECHNICAL SCIENCES

LONG RANGE FORECASTING OF RIVER FLOWS

Sevgi Cavdar
Ph.D., Assistant Professor, Department of Civil Engineering, Faculty of Engineering, Sivas Cumhuriyet University, Sivas, Turkey
Hikmet Kerem Cigizoglu
Ph.D., Retired Professor, Civil Engineering Department, Civil Engineering Faculty, Istanbul Technical University, Istanbul, Turkey
https://doi.org/10.5281/zenodo.18631627

Abstract

A study focused on long range forecasting of river flows is presented. Two methods, artificial neural networks (ANNs) and multi linear regression (MLR) are employed for forecasting with time ranges changing between 1 day and 17 days. It was seen that the MSE values of ANN for the first 4 time ranges (1, 2 , 3 and 4 days) are lower compared with the corresponding MLR ones. For 5 days ahead and further forecasting the MSE values of MLR are lower than the ANN ones. The R^2 values for the time ranges 1, 2, 3, 4 and 5 days are equal for both methods. The MLR method has higher R^2 values for time ranges starting with 6 days.

Keywords: River flow; artificial neural networks; multilinear regression; long range forecasting.

Introduction

River flow forecasting is quite important for hydrology. In hydrological problems like flood modeling and dam spillway modeling river flow forecasting plays a significant role. In the literature the forecasting studies are based on one step ahead forecasting [1-15] . If the inputs of the forecasting model covers the time interval 't-1'-'t-n' the corresponding output time is 't'. For example if the time step is 'day' the available forecasting works are concentrated on 'one day ahead' forecasting. The long range forecasting, i.e. forecasting for output times 't+1',....,'t+n' is not examined in the literature. The long range forecasting can shed light to the future of the hydrological studies. If the researcher or engineer can have information for the future times the required precautions and adjustments on the hydrological works can be taken in advance. In the presented study different forecasting studies are employed for the long range forecasting of the selected river flow time series.

The daily mean flow data employed in the study is provided by the web site of United States Geological Survey (USGS). The measurement station name is Rio Grande at Otowi Bridge, NM. The station ID is 08313000. The basin area is 14300 sq.mi. The latitude and longitude of the station are 355229 and 1060830, respectively. The data record covers the time interval between 1955 and 1989. The training period for the forecasting is the first 80% of the whole record whereas the testing time period consists of the following 20% of the data. The basic statistics for the whole record, training record and testing record are presented in Table 1. The mean and standard deviation values of the testing period are higher compared with the training period (Table 1). The skewness of the training data is higher than the testing data (Table 1). The skewness values point to the positively skewed marginal probability distribution for the daily flow record. The first- and second-order autocorrelations are quite close to 1 pointing to the strong internal dependence of the data (Table 1).

Data description

Table 1. The description of the data statistics.

	Whole data	Training data	Testing data
Mean, x_m, (m^3/s)	39.18	34.54	57.75
Standard deviation, s_x, (m^3/s)	42.43	37.12	55.30
Skewness, c_{sx}	2.73	2.95	2.0
1st order autocorrelation, r_1	0.99	0.988	0.994
2nd order autocorrelation, r_2	0.978	0.974	0.985

The forecasting results

In the presented study two forecasting methods are used. The first one is an artificial neural network (ANN) method, the feed forward back propagation method (FFBP). The details of this method is available in the ANN literature [3-7]. The second method is a statistical method, i.e.the multi linear regression (MLR) method. For the ANN simulations the ANN toolbox of the MATLAB software is employed. The statistics toolbox of the EXCEL was used for the MLR computations. Two statistics are selected for the evaluation of the testing stage forecasting results, i.e. the mean square error (MSE) and the determination coefficient (R^2).

The input range of the ANN architecture constituted of 10 previous daily flows (times: t-10, t-9, t-8, t-7, t-6, t-5, t-4, t-3, t-2, t-1). The ANN had one hidden layer. The unique output of the ANN output layer corresponds to the future daily flow (time: t, t+1, t+2,........, or t+16). Similarly the MLR had 10 independent variables representing 10 previous daily flows and one dependent variable for the future time range. The MSE and R^2 statistics for the testing data for 17 different future time ranges are presented in Table 2.

Table 2. Testing data forecasting statistics for ANN simulations and MLR for different time ranges.

Time range (days)	**MSE (m^6/s^2) (ANN)**	**MSE (m^6/s^2) (MLR)**	**R^2 (ANN)**	**R^2 (MLR)**
1	27	30	0.99	0.99
2	89	92	0.97	0.97
3	168	169	0.94	0.94
4	247	259	0.92	0.92
5	428	355	0.88	0.88
6	442	449	0.84	0.85
7	598	542	0.81	0.82
8	673	633	0.78	0.79
9	871	720	0.72	0.76
10	921	805	0.71	0.74
11	1082	885	0.66	0.71
12	1195	961	0.63	0.69
13	1338	1033	0.59	0.66
14	1311	1101	0.59	0.64
15	1377	1162	0.57	0.62
16	1465	1220	0.54	0.61
17	2198	1272	0.4	0.59

The MSE values of ANN for time ranges 1, 2 , 3 and 4 days are lower compared with the corresponding MLR ones (Table 2). Starting with the time range 5 (5 days ahead, t+4) the MSE values of MLR become lower than the ANN ones. The R^2 values for the time ranges 1, 2, 3, 4 and 5 days are equal for both methods. The deviation between R^2 values of two methods start from time range 6 on in the favour of MLR method. The lowest R2 values are obtained for time range 17 days (t+16) for the testing stage (R^2=0.40 for ANN, R^2=0.59 for MLR, Table 2). The variation of MSE and R^2 values with respect to different time ranges for both the ANN and MLR methods are shown in Figs. 1 and 2.

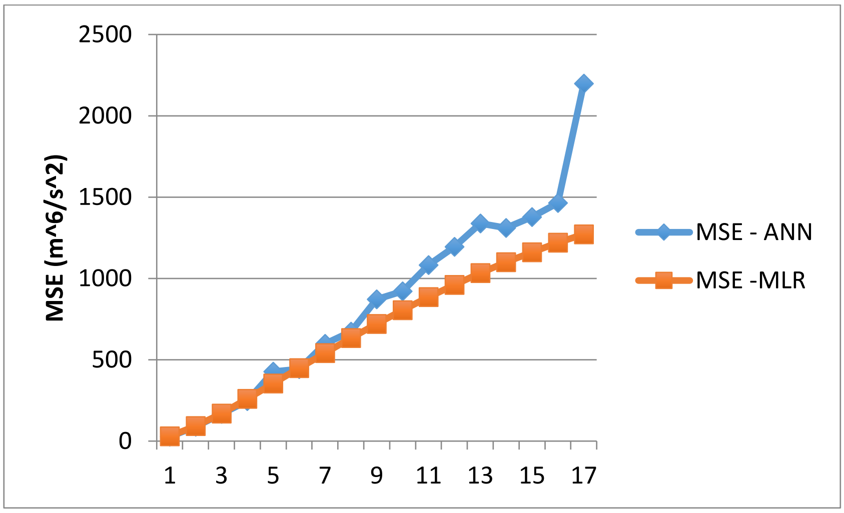

Fig.1. *The variation of MSE values with respect to time range (days) for ANN and MLR methods.*

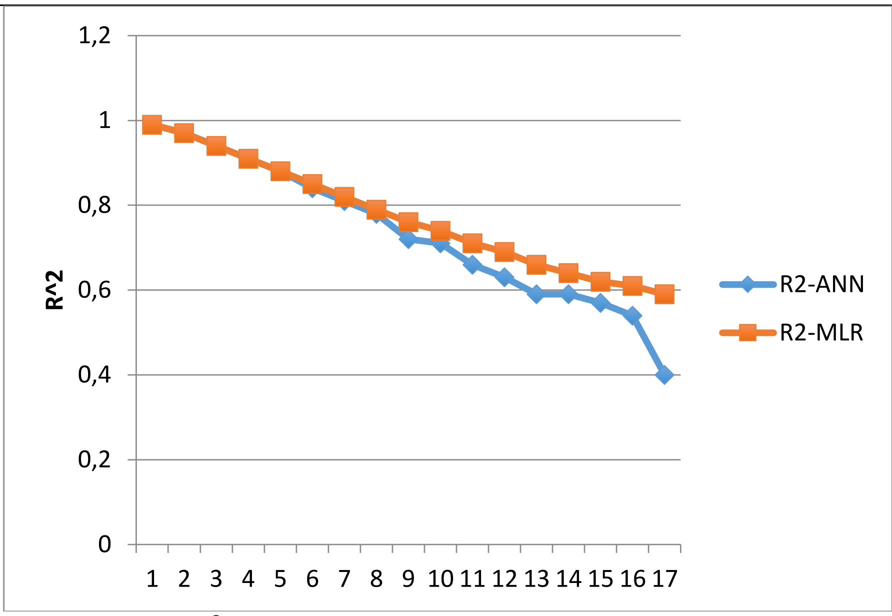

Fig. 2. *The variation of R^2 values with respect to time range (days) for ANN and MLR methods.*

The graphics for the scatter plots obtained with both methods for time ranges 1, 4 and 10 days (t, t+3 and t+9) are plotted in Figs. 3, 4, 5, 6, 7, 8.

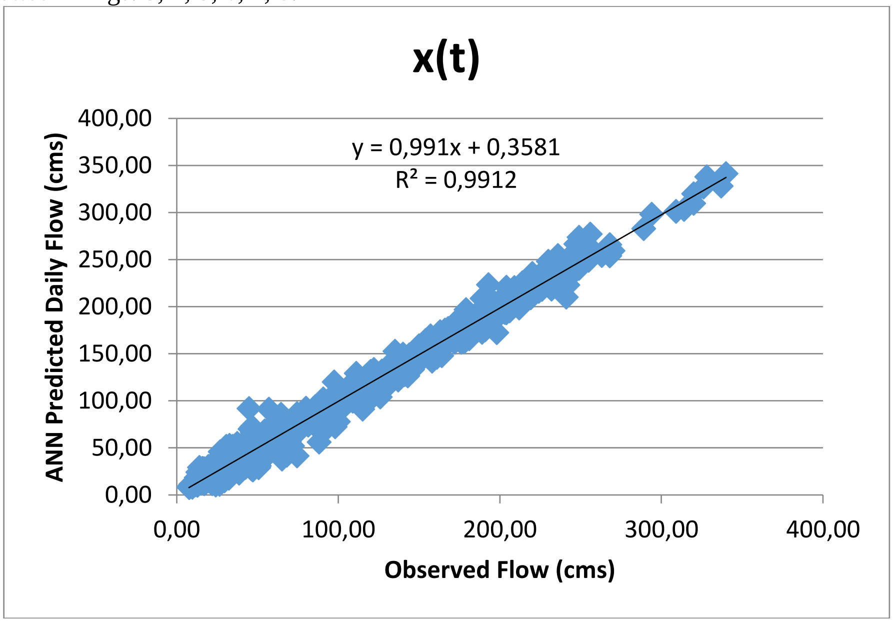

Fig.3. *The scatter plot for ANN predictions for time range 1 (1 day).*

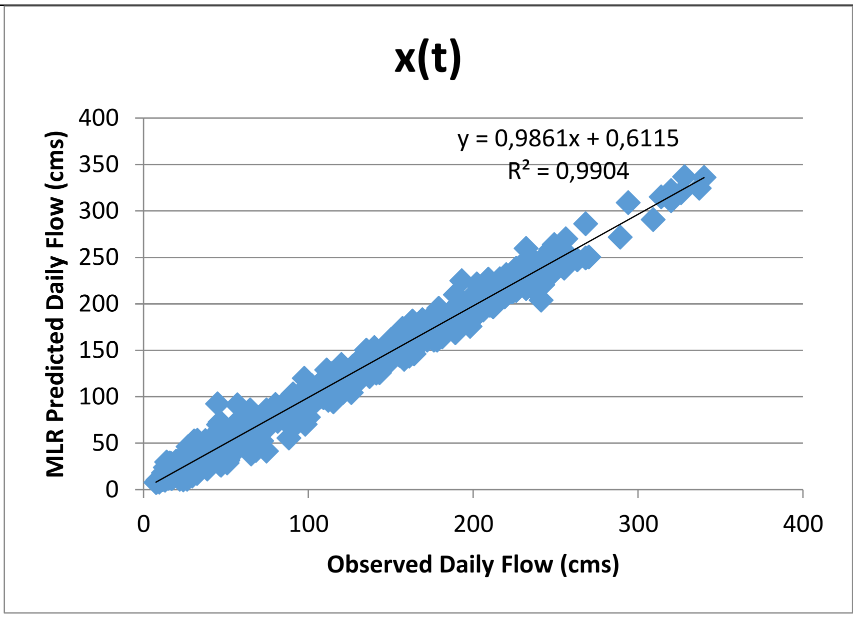

Fig.4. *The scatter plot for MLR predictions for time range 1 (1 day).*

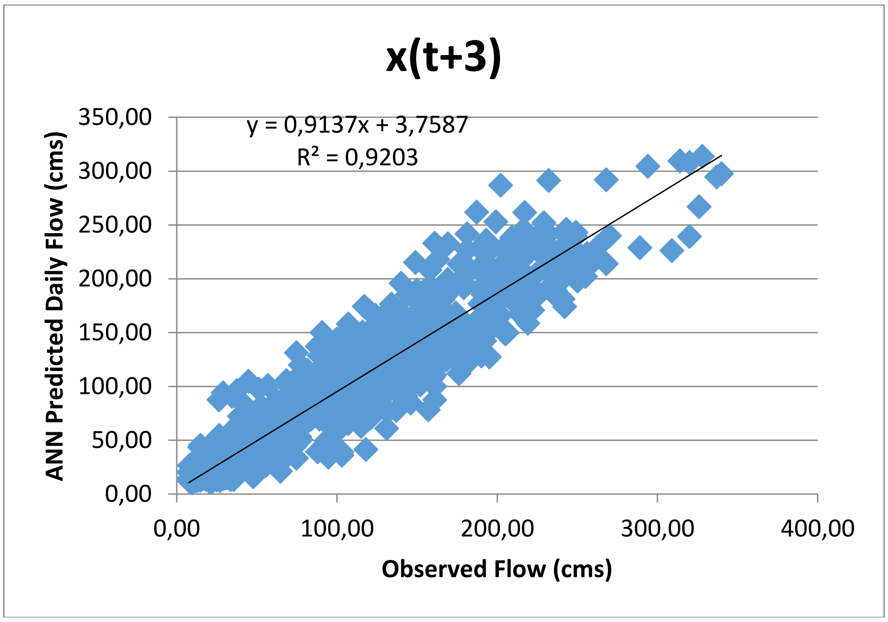

Fig.5. *The scatter plot for ANN predictions for time range 4 (4 days).*

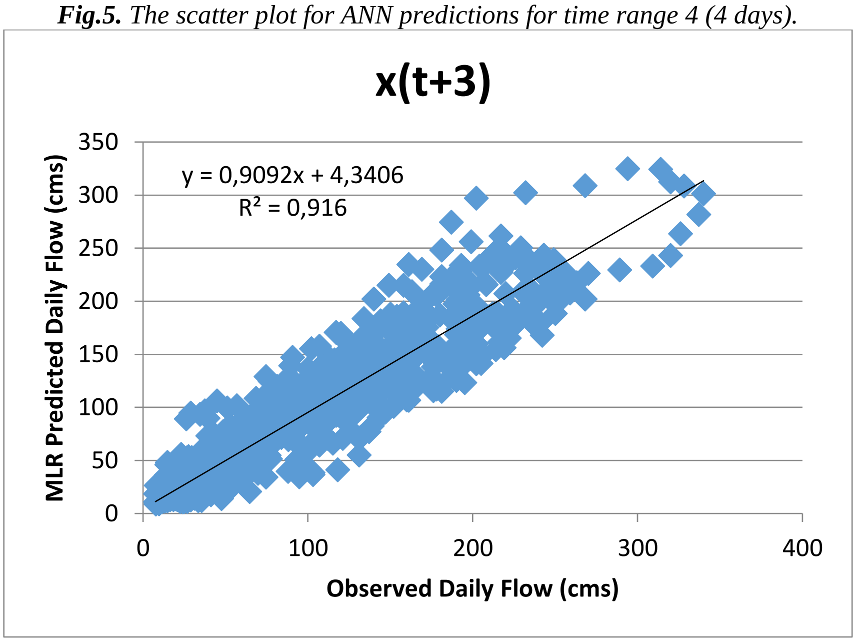

Fig.6. *The scatter plot for MLR predictions for time range 4 (4 days).*

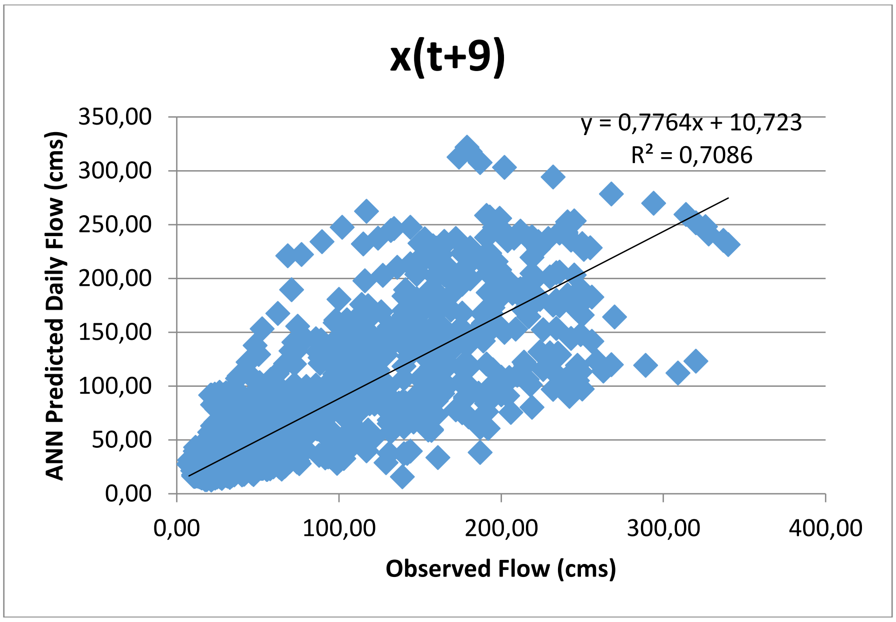

Fig.7. *The scatter plot for ANN predictions for time range 10 (10 days).*

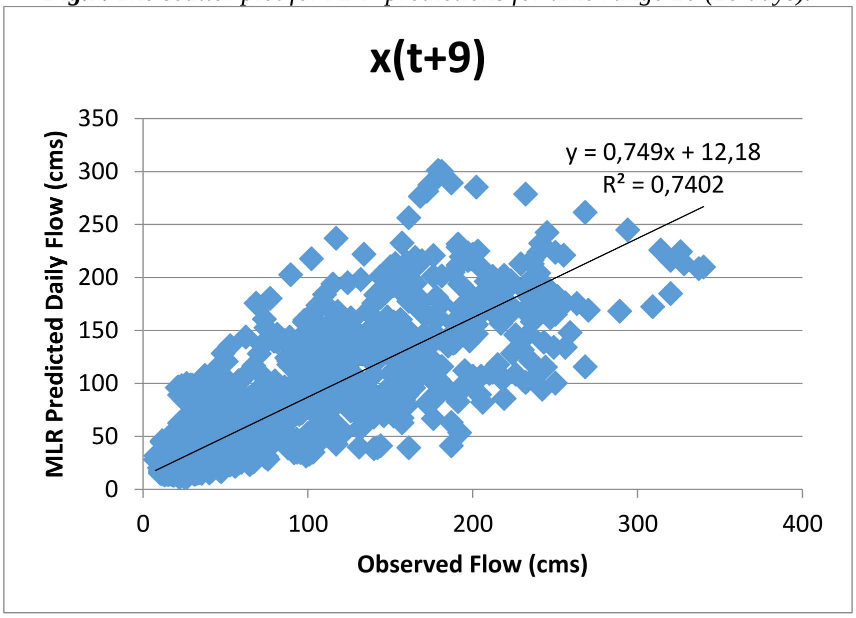

Fig.8. *The scatter plot for MLR predictions for time range 10 (10 days).*

In the case with time range 1 day the scattering is quite limited for both ANN and MLR methods with R^2 values nearly equal to 1 (Figs. 3 and 4). As time range increases to 4 days the scattering becomes obvious with R^2=0.92 for both methods (Figs. 5 and 6). The time range 10 days shows noticeable dispersion with slightly higher R^2 for MLR (Figs. 7 and 8).

Discussion and conclusion

The presented study examined the performance of two methods, i.e.ANN and MLR, for long range forecasting. The time range varied between 1 day and 17 days. It was seen that the MSE values of ANN for the first 4 time ranges (1, 2 , 3 and 4 days) are lower compared with the corresponding MLR ones. For 5 days ahead and further forecasting the MSE values of MLR are lower than the ANN ones. The R^2 values for the time ranges 1, 2, 3, 4 and 5 days are equal for both methods. The MLR method has higher R^2 values for time ranges starting with 6 days. The long range forecasting study can be considered as a beneficial tool in water resources studies like dam spillway management. As well known the dam spillway is one of the most significant parts of the dam structure. The performance of the spillway can be controlled if information is available for the future flows in the coming time intervals (days in the presented study). The potential damage on such a dam structure due to an excessive increase on the entering flow can be prevented or decreased with an effective forecasting strategy.

References:

1. Alp, M., Cigizoglu, H.K. (2007).“Suspended sediment estimation by feed forward back propagation method using hydro meteorological data”, Environmental Modelling & Software, 22(1), 2-13.

2. ASCE. (2000). “Task committee on applicaton of artificial neural networks in hydrology: Artificial neural networks in hydrology. Part II—Hydrologic applications.” J. Hydr. Eng., 10.1061/(ASCE)1084-0699 (2000)5:2(124), 124–137.

3. Cigizoglu, H.K. (2003). ”Estimation, forecasting and extrapolation of river flows by artificial neural networks.”, Hydrological Sciences Journal, 48(3), 349-361.

4. Cigizoglu, H.K.(2003).”Incorporation of ARMA models into flow forecasting by artificial neural networks.”, Environmetrics, 14, 4, 417-427.

5. Cigizoglu, H.K.(2004).”Estimation and forecasting of daily suspended sediment data by multi layer perceptrons.”Advances in Water Resources, 27, 185-195.

6. Cigizoglu, H.K. (2005).“Application of the Generalized Regression Neural Networks to Intermittent Flow Forecasting and Estimation”, J. Hydr.Eng., 10(4), 336-341.

7. Cigizoglu, H.K.(2005). “Generalized Regression Neural Network in Monthly Flow Forecasting”, Civil Engineering and Environmental Systems, 22 (2), 71-84.

8. Cigizoglu, H.K., Alp, M. (2006).“Generalized Regression Neural Network in modelling river sediment yield.”, Advances in Engineering Software, 37(2), 63-68.

9. Kişi, Ö. (2004).“River flow modeling using artificial neural networks.” J. Hydr. Eng., 10.1061/(ASCE)1084-0699(2004)9:1(60),60–63.

10. Kişi, Ö. (2007).“Streamflow forecasting using different artificial neural network algorithms.” J. Hydrol. Eng.,10.1061/(ASCE)1084-0699 (2007)12:5(532), 532–539.

11. Linares-Rodriguez, A., Lara-Fanego, V., David Pozo-Vazquez, D., Tovar-Pescador, J. (2015).” One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model.” J. Hydr.Eng., 20(9): 05015001.

12. Sertel, E., Cigizoglu, H.K., Sanli, D. U. (2008).“Estimating daily mean sea level heights using artificial neural networks.”, Journal of Coastal Research, 24(3), 727–734.

13. Tayfur, G. (2012). Soft computing in water resources engineering: Artificial neural networks, fuzzy logic and genetic algorithms, WIT Press, Southampton, U.K.

14. Ustaoglu, B., Cigizoglu, H.K., Karaca, M.(2008).”Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods.”, Meteorological Applications, 15 (4), 431-445.

15. Wu, J., Han, J., Annambhotla, S., and Bryant, S. (2005). “Artificial neural networks for forecasting watershed runoff and stream flows.” J. Hydr.Eng., 10.1061/(ASCE)1084-0699(2005)10:3(216), 216–222.