KSCE Journal of Civil Engineering (2016) 20(2):971-977

DOI 10.1007/s12205-015-0298-z

Water Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

A Comparison of Artificial Neural Networks (ANN) and Local Linear Regression (LLR) Techniques for Predicting Monthly Reservoir Levels

Muhammad Ali Shamim*, Muhammad Hassan, Sameel Ahmad***, and Muhammad Zeeshan******

Received June 2, 2014/Revised November 15, 2014/Accepted February 7, 2015/Published Online June 19, 2015

Abstract

Storage dams play a very important role in irrigation especially during lean periods. For proper regulation one should make sure the availability of water according to needs and requirements. Normally regression techniques are used for the estimation of a reservoir level but this study was aimed to account for a non-linear change and variability of natural data by using Gamma Test, for input combination and data length selection, in conjunction with Artificial Neural Networking (ANN) and Local Linear Regression (LLR) based models for monthly reservoir level prediction. *R*esults from both training and validation phase clearly indicate the usefulness of both ANN and LLR based prediction techniques for Water Management in general and reservoir level forecasting in particular, with LLR outperforming the ANN based model with relatively higher values of Nash-Sutcliffe model efficiency coefficnet (R^2) and lower values of Root Mean Squared Error (RMSE) and Mean Biased Error (MBE). The study also demonstrates how Gamma test can be effectively used to determine the ideal input combination for data driven model development.

Keywords: *reservoir level, Artificial Neural Network ANN, Local Linear Regression LLR, Gamma Test (GT)*

1. Introduction

Storage dams play a vital role for effective water resource management, especially in arid area where abundant amount of water is only available for a certain part of the year. With increase in population and rapid industrialization and development, the world's water resources are being stretched. As per UNESCO's World Water Development Report (2003), a 30-40% decrease in the available quantity of water is predicted in the next 20 years, which is alarming. As far as Pakistan is concerned, the per capita water availability is quite close to the threshold of being declared a water scarce country (WAPDA, 2010). Along with this, recent extreme events within the country (extreme rain and floods in 2010 and 2011 and prolong drought in 2009) clearly indicate that effective management of water reserves is the need of the hour. Storage dams are the main source of water conservation within Pakistan. Therefore, accurate prediction of monthly water levels, within a reservoir is very important for an estimation of available water storage. These are also important for developing an efficient reservoir operation policy, so that water releases can be rationalized depending upon need and availability.

As far as prediction of reservoir water levels is concerned, two approaches are currently prevailing. One is based on annual reservoir volume model, water balance or water budget. The second one is based on statistical water balance equation, in order to generate seasonal volume on the basis of seasonal climatic variables like precipitation, evaporation and runoff (Güldal and Tongal, 2010). Nowadays, owing to technological advancement, Artificial Intelligence (AI) based techniques are becoming increasingly popular in prediction of hydrological processes (Shamim *et al.*, 2010). These include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Fuzzy Logic (FL), Adaptive Neuro Fuzzy Inference Systems (ANFIS), Decision Trees (DT) or a hybrid combination of these techniques (Remesan *et al.*, 2010) with ANNs being the most popular. A number of ANN structures and topologies are being used for hydrological and meteorological prediction, among which the Feed Forward Neural Networks are the most popular ones (Güldal and Tongal, 2010).

Over the years, ANN based models have been used for streamflow prediction (Han *et al.*, 2002, Han *et al.*, 2007, Remesan *et al.*, 2009; Remesan *et al.*, 2010), reservoir inflow prediction (Hassan *et al.*, 2014), rainfall estimation (Kissi and Cimen, 2012; Alqudah *et al.*, 2013), Solar radiation estimation (Remesan *et al.*, 2008; Shamim *et al.*, 2010; Shamim *et al.*, 2014), Land use classification (Srivastva *et al.*, 2012) and wind

*Assistant Professor, Dept. of Civil Engineering Bursa Orhangazi University, Bursa 16310 Turkey (Corresponding Author, E-mail: muhammad.shamim@bou.edu.tr)
**Lecturer, Dept. of Civil Engineering Mirpur University of Science and Technology, Mirpur 10250, Pakistan (E-mail: hassan25.arif@ gmail.com)
***Research Scholar, Dept. of Civil Engineering, University of Engineering and Technology Taxila, 47050, Pakistan (E-mail: smaeelahmad@gmail.com)
****Research Scholar, Dept. of Civil Engineering, University of Engineering and Technology Taxila, 47050, Pakistan (E-mail: mzeeshan.21@gmail.com)

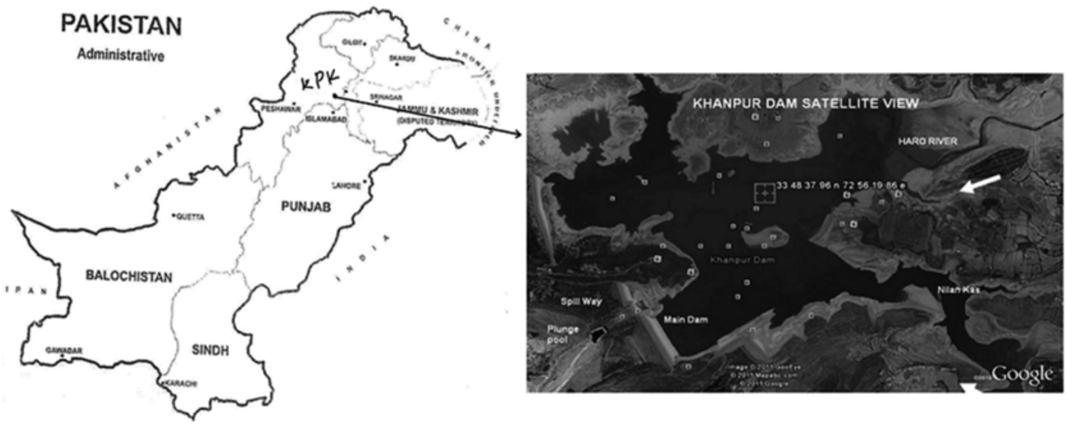

Fig. 1. KhanPur Dam Ariel View (Source Google Map)

speed modelling (Ishak *et al.*, 2013), modelling snowmelt–runoff (Matheussen and Thorolfsson 1999; Tokar and Markus 2000), sediment transport prediction (Tayfur and Güldal, 2006), groundwater modeling (Rogers and Dowla, 1994; Lallahem and Mania, 2003; Shamim *et al.*, 2004) ecological response to climate change (Trigo and Palutikof, 1999) and reservoir operation (Hasebe and Nagayama, 2002; Jain *et al.*, 1999b; Raman and Chandramouli, 1996). American Society of Civil Engineering (ASCE's) task committee has also acknowledged the importance of ANNs as an important forecasting tool as elaborated in ASCE (2000a and b).

Local Linear Regression (LLR) is a considered to be an alternative technique to ANN for performing nonlinear time series prediction (Kemp *et al.*, 2006). Over the years it has increasingly been used in hydrological prediction, examples of which include Solar Radiation estimation (Remesan *et al.*, 2008a; Ahmadi *et al.*, 2009; Shamim *et al.*, 2010; Shamim *et al.*, 2014) and Streamflow prediction (Remesan *et al.*, 2008b; Remesan *et al.*, 2009; Hassan *et al.*, 2014; Jajarmizadeh *et al.*, 2014), but to date it has not been used for reservoir level prediction. This study will evaluate the accuracy of LLR based non parametric technique for predicting monthly reservoir levels in comparison with ANN, which is considered to be a recognized methodology for hydrometeorological prediction. The study is novel in the sense that it is the first time that ANN and LLR based data driven techniques are used for mean monthly reservoir level prediction, for efficient water management within Pakistan. The methodology is applied to a Khanpur Dam, located in the north of Pakistan, mainly used for irrigation and water supply. This study is also novel in the sense that it incorporates the use of Gamma Test for the selection of best input combination, a criterion not previously considered for reservoir level prediction. As far as the data for the study is concerned, current and up to two previous months' mean records (t, t–1 and t–2) for Inflows, Outflow, Precipitation, Evaporation and Seepage were used to predict month ahead mean reservoir water levels. Out of a total of 2^n-1 combina- tions, five best were selected and a total of 15 models (03 for each combination) were developed. Output results from both training and validation were compared with observed data in order to determine which set of input combination and model structure leads to the most accurate prediction of future water levels for the selected reservoir.

2. Methodology

2.1 Study Area and Data Sets

Khanpur dam, located on Haro River near the town of Khanpur, Khyber Pakhtunkhuwa province Pakistan was selected as the study area. It is located at a distance of about 40 km from the capital Islamabad. Khanpur Dam (Fig. 1) is basically a storage dam that provides water for irrigation as well as for domestic and industrial use in periods of high demands to adjacent areas. Two canals originate from the dam; Right Bank Canal and Left Bank Canal to irrigate an area of 76.04 km^2 in Khyber PK and 71.54 km^2 in the Punjab Province with discharge capacity of 15.5 m^3/s. Datasets comprising of mean monthly inflows, outflow, rainfall, evaporation, seepage and reservoir levels for a period of 17 years from 1988 to 2004 were collected from Water and Power Development Authority (WAPDA) which is responsible for the operation and maintenance of Khanpur Dam. Monthly datasets from 1988 to 2000 were used for model training and 2001-2004 for model validation. As far as the selection of inputs is concerned, current and upto two previous months' observational records were considered so as to identify the best input combination with the aid of Gamma test

Table 1. Inputs Used for Model Development

Sr. No.	Input	Sr. No.	Input
1	Inflow (t)	9	Rainfall (t-2)
2	Inflow (t-1)	10	Evaporation (t)
3	Inflow (t-2)	11	Evaporation (t-1)
4	Outflow (t)	12	Evaporation (t-2)
5	Outflow (t-1)	13	Seepage (t)
6	Outflow (t-2)	14	Seepage (t-1)
7	Rainfall (t)	15	Seepage (t-2)
8	Rainfall (t-1)		

*Where t represents time (Present Month), t-1 (Previous Month), t-2 (Month before previous)

(Table 1).

2.2 Gamma Test, V-ratio and M-Test

Developed by Agalbjörn *et al.* (1997) in order to perform nonlinear analysis, Gamma Test (GT) facilitates us to estimate the best mean square error on the measured output data which may be attained by a smooth and noiseless model for a given set of real variables. Further elaboration about GT can be found in Končar (1997); Agalbjörn *et al.* (1997); Chuzhanova *et al.* (1998); Tsui (1999); Tsui *et al.* (2002); Durrant (2001) and Jones *et al.* (2002).

Suppose an initial dataset $\{(\boldsymbol{x}_i, y_i), 1 \le i \le M\}$ consists of a vector $\boldsymbol{x}$ input confined to some close delimited set $C \subset R^m$ and, without loss of generalization, corresponding scalar output y ($y \subset R^m$). In fact this initial dataset is used to develop an algorithmic program which is capable to predict the output y by using associated input vector $\boldsymbol{x}$. The output y depends upon the input $\boldsymbol{x}$ i.e. $\boldsymbol{x}$ influences y. Assume the output y is composed of a smooth function and noise r then mathematically,

$$y = f(\boldsymbol{x}_1 + \cdots\cdots\cdots \boldsymbol{x}_m) + r \tag{1}$$

Random variable 'r' may be defined as the noisy part of the noiseless output. It can be assumed that the mean of the stochastic variable r is zero due to the fact that a constant systematic error can be engaged into an unknown function f to limit the variance. For any periodic input data set $T = \{\boldsymbol{x}_1, \cdots\cdots, \boldsymbol{x}_m\}$, the first and second nearest neighbors of $\boldsymbol{x}_i$ are those points having minimum distances from $\boldsymbol{x}_i$ and minimum indexes from j_1 and j_2, respectively. Now generally in the set T, the kth nearest neighbor of each vector $\boldsymbol{x}_i$ can be denoted as $\boldsymbol{x}_N[i, k]$. Similarly the corresponding value of the output will be $y_N[i, k]$

Gamma test is based upon the delta function of input vectors

$$\delta_m(k) = \frac{1}{M}\Sigma_{i=1}^{M}|\boldsymbol{x}_{N[i,k]} - \boldsymbol{x}_i|^2 \tag{2}$$

$$\gamma_m(k) = \frac{1}{2M}\Sigma_{i=1}^{M}|y_{N[i,k]} - y_i|^2 \tag{3}$$

For $1 \le k \le p$, considering p as a fixed integer, the pairs $[\delta_m(k), \gamma_m(k)]$ were computed and a linear regression equation was developed so as to determine

$$\text{Gamma } (\gamma) = A\delta + \Gamma \tag{4}$$

The vertical intercept $\Gamma(\delta_M(k) = 0)$ is the measure of variance and is very suitable in determining the best mean square error graphically. Slope A determines the model complexity; steeper the gradient higher will be the model complexity (Evans and Jones, 2002). Results within GT can be standardised by considering a scale variant noise estimate, normally lying between zero and 1, called V_{ratio} and define as

$$V_{\text{ratio}} = \frac{r}{\sigma^2(y)} \tag{5}$$

Variance of output 'y' is represented by $\sigma^2(y)$. A value close to zero indicates the high degree of predictability of output 'y' from selected inputs.

M-test, is used to decide the length of the dataset to produce a stable asymptote. It is basically a mathematical criterion that weathers an infinite series of functions converges uniformly or not. Value of gamma statistics is determined by using increasing number of points M. Number of points M or data length is selected in a way that it stabilizes the model for a specific value of gamma statistic Γ (which is the best estimate of mean square error for the development of a smooth model). This test enables us to overcome the over fitting problems in the nonlinear modeling.

The above mentioned tests were used for selection of best input combination and data length that will consequently be used for model development for reservoir level forecasting.

2.3 Modelling Techniques

2.3.1 Local Linear Regression (LLR)

LLR is a non-parametric technique used for low dimensional estimations. Even for small input data modeling it gives quite satisfactorily and reliable results. One of the significant advantages of using LLR is that it can produce reasonable reliable model output locally with relatively small amount of sample data (Remesan *et al.*, 2008). Therefore, for regions with high data density, much accurate predictions can be produced. Initially three data point are required to initiate the prediction; the newly developed points are then used for further prediction process.

LLR performs linear regression using a set of p_{max} nearest points to a query point that in turn leads to the production of a linear model, located in the vicinity query point. Computation of p_{max} nearest points to the query point is achieved using *kd*-tree algorithm, a fast nearest neighbor search algorithm, elaborated in detail in Kemp (2006).

A linear matrix equation ($Xm = Y$) is solved for the unknown 'm' where 'X' is a $p_{max} \times d$ matrix of p_{max} input points in d-dimension, x_i ($1 \le i \le p_{max}$) are the nearest neighbor points while 'y' represents a column vector having a length equal to p_{max}. Also, 'm' represents a column vector of parameters that provides optimal mapping from 'X' to 'y' using (Eq. 6)

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1d} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2d} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{xpmax1} & x_{xpmax2} & x_{xpmax3} & \cdots & x_{xpmaxd} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_d \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{max} \end{bmatrix} \tag{6}$$

If X is nonsingular square matrix then the above linear matrix can be solved by an equation; $m = X^{-1} \times Y$, If it is a non-square singular matrix then a vector m should be found which minimizes the $|Xm - y|^2$ and a unique solution can be obtained by a linear equation ($m = X^{\#}y$) where $X^{\#}$ is a pseudo inverse matrix (Penrose, 1955, 1956). Further details about LLR can also be found in Kemp (2006).

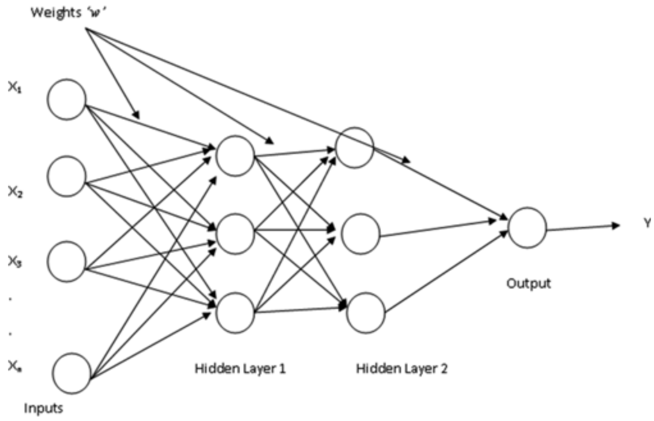

Fig. 2. ANN Conceptual Framework (2-hidden layer structure)

2.3.2 Artificial Neural Networking Technique (ANNs)

Artificial neural networking technique is in fact inspired by the neural system of human brain as a specific act is associated with a specific reason. The reason and its respected act is captured and then related by a complex structure of biological neurons. ANNs involves a similar neural networking intelligent system that captures the composite relationships between inputs and outputs. Initially developed by McCulloch and Pitts (1943) and later on modified by Rosenblatt (1962), ANNs normally consist of three layers; an input layer, a hidden layer (processing unit) and an output layer (Fig. 2). The use of one hidden layer or more than one depends upon the type of ANNs being used and in this study two hidden layers were used, utilizing hit and trial approach. Now consider ANNs as a processing unit in which we introduce the inputs and get outputs after processing by a nonlinear neural approach. If $x_1, x_2, x_3 ..., x_m$ are the inputs which are sensitive and has an influence on the predicted values of $y_1, y_2, y_3 ..., y_m$ then.

If we explore the processing unit, the input nodes and the nodes in the hidden layer/layers interact with each other in a way that the nodes in a same layer does not have any connection but the nodes of different layers interact with each other and have some connection strength or weight. The assigning of connection strength to each node with the other node is done in a way that it gives value equal to or nearly equal to the value of desired output.

The process requires a considerable amount of historical data of inputs as well as of outputs to perform training phase. Greater the uncertainty, greater, will be the data required for the development of a smooth and reliable model. Once training phase is performed and the weights are assigned, errors (both systematic and random) are computed to check the accuracy of developed equation/model. Model training is normally carried out with the aid of learning (training) algorithms like the Back Propagation training algorithm, quasi-Newton BFGS training algorithm, Conjugate Gradient (CG) and the Levernberg-Marquardt (LM) training algorithm etc. As far as this study is concerned, model training for artificial neural network has been carried out using Back Propagation algorithm and quasi-Newton type Broyden–Fletcher–Goldfarb–Shanno (BFGS) training algorithm. Model results have also been compared with Local Linear Regression (LLR) based model. Another important criterion that has a significant effect on model output is the choice of hidden layers for model training. Minsky and Papert (1969) and Jones (2004) have highlighted the importance of incorporating two hidden layers, for model development and the same has been incorporated in this study.

3. Results

3.1 Model Input Combination and Data Length Selection

Gamma Test (GT) and M-test were utilized for the determination of best input combination and data length so as to build a smooth and reliable model. All inputs employed for determining the best combination are elaborated in Table 1. A total of 500 combinations were explored to determine their influence on mean monthly reservoir level by computing their Gamma statistic Γ. In general, 2^{n-1} input combinations are possible but only the most realistic (500) were formed. Variation of gamma statistic Γ for all 500 input combinations (experiments) is shown in Fig. 3.

Out of the above mentioned 500 combinations, five best combinations (experiments) were consequently selected to determine the data length (M) required to produce a smooth model. For each combination, number of nearest neighbors was limited to the default value of 10. Gamma test and M-test statistics for these combinations are elaborated in Table 2. A mask is an artificial layer that defines a particular combination by including or excluding specific input or inputs in that combination. ‘1’ denotes inclusion and ‘0’ denotes exclusion of a particular input.

3.2 Modeling Results

The selected combinations were used to build ANN and LLR based models for monthly reservoir level prediction. Two types of feed forward ANN models were trained that were based upon

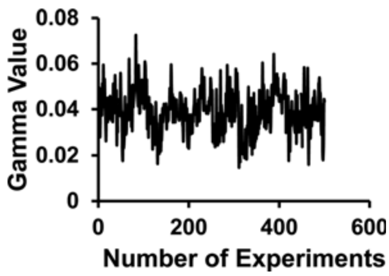

Fig. 3. Variation of Gamma Statistic with Number of Input Combinations (experiments)

Table 2. Gamma Test and M-test Results for Selected Input Combinations

Test	Mask*	Gamma Stastic	M-test (vectors)
1	111100111111111	0.024	60
2	111110100101111	0.016	135
3	111110101111101	0.017	120
4	111110100111111	0.015	135
5	110111110100111	0.029	85

‘0’ denotes exclusion and ‘1’ denotes inclusion of an input

Table 3. LLR, BFGS and BP Model Structure for Monthly Reservoir Level Prediction

Test	LLR		BFGS			Back Propagation		
	Nearest Neighbors	Gamma (Γ)	Nodes (layer 1)	Nodes (layer 2)	Gamma (Γ)	Nodes (layer 1)	Nodes (layer 2)	Gamma (Γ)
1	15	1.13E-05	5	4	1.13E-05	5	4	1.39E-05
2	15	1.01E-05	5	4	1.17E-05	5	4	1.67E-05
3	15	1.02E-05	5	4	1.93E-05	5	4	1.81E-05
4	15	1.09E-05	5	4	1.63E-05	5	4	1.51E-05
5	15	9.65E-06	5	4	1.45E-05	5	4	1.41E-05

two layered Broyden Fletcher Goldfarb Shanno (BFGS) and Back Propagation (BP) Training Algorithm. Results from these models were also compared with LLR based models, within the winGamma Environment. For LLR the selection of Nearest Neighbors (NN) was made by trial and error method that gave least gamma value, which was found to be 15 for this study (Table 3). Similarly both ANN models were provided with two hidden layers with the number of neurons in each layer selected through a trial and error procedure, which were found to be '5' and '4' respectively for the first and second hidden layer (Table 3).

Scatter plots for all developed models were drawn (Figs. 4, 5 and 6) and coefficient of determination (R-square) values calculated. Additionally, for comparison with other models, statistical parameters like Root Mean Square Error (RMSE) and Mean Bias Error (MBE) were also computed along with R-square (Table 4(a) and (b)).

As far as the modeling results are concerned, Gamma test (Table 2) indicated that input combination 4 (1111101001111110000) and 2 (1111101001011110000) will lead to relatively accurate estimation of mean monthly reservoir estimation, as their Gamma Statistic values were relatively low as compared to the rest. The same has been verified from results of Tables 4(a) and (b) as both these combinations perform reasonably well for training portion and best for validation part. For the validation part, for LLR based modeling, combination no. 4 had the best value of R-square (0.91) and MBE (0.47 m) and 3rd best value for RMSE (1.96 m), as elaborated in Table 4(b). On the other hand for ANN based BFGS and BPNN models, combination no. 2 had the best values for R-square (0.90 and 0.88 respectively) and RMSE (1.91 m and 1.83 m respectively) and relatively lower values for MBE (0.85 m and 0.05 m) respectively (Table 4(b)). Therefore, it can rightly be concluded that combinations 2 and 4 outperformed the rest of the input

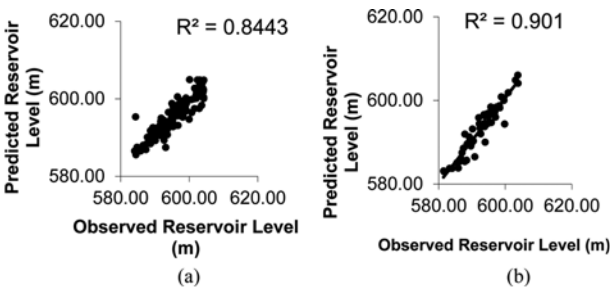

Fig. 4. Scatter Plots for Best LLR model for Mean Monthly Reservoir Level Prediction: (a) Training, (b) Validation

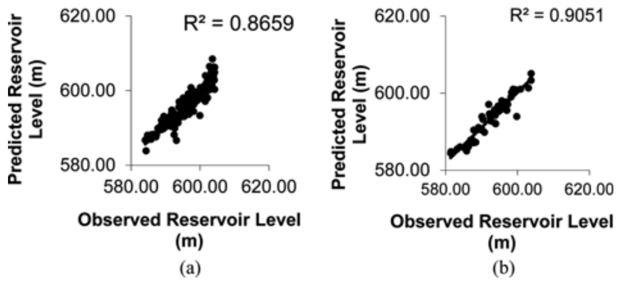

Fig. 5. Scatter Plots for Best BFGS Based ANN Model for Mean Monthly Reservoir Level Prediction: (a) Training, (b) Validation

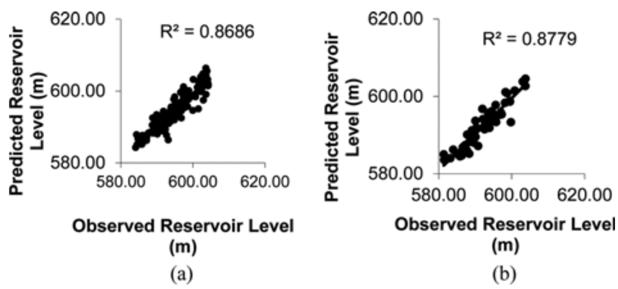

Fig. 6. Scatter Plots for Best Backpropagation Based ANN Model for Mean Monthly Reservoir Level Prediction: (a) Training, (b) Validation

combinations, for predicting mean monthly reservoir levels at Khanpur dam, with combination 2 the best amongst ANN based models and combination no 4, being the best amongst LLR based models. As far as overall performance is concerned, LLR based model outperformed ANN with relatively higher values of R^2 and lower values of RMSE and MBE, respectively. Scatter plots for best modeling results for each of the combinations are shown in Figs. 4, 5 and 6 respectively.

4. Conclusions

The study was focused on the estimation of mean monthly reservoir levels, crucial in the reservoir management. Gamma test and M-test were used for identification of best input combination and data length selection for month ahead reservoir level prediction. A total of 500 input combinations were analyzed out of which best 5 were selected for model development on the basis of best (least) value of Gamma statistic that would give best (least) mean square error estimate of modeling results. M-test

Table 4. Comparative Performance of LLR, BFGS (ANN) and BP (ANN) Models for Monthly Mean Reservoir Level Prediction: (a) Training, (b) Testing Period

(a)

Model	LLR			BFGS (NN)			Back Propagation (NN)		
Combination	RMSE (m)	MBE (m)	R^2	RMSE (m)	MBE (m)	R^2	RMSE (m)	MBE (m)	R^2
1	2.21	-0.10	0.83	2.18	-0.05	0.83	2.06	0.02	0.85
2	2.22	-0.03	0.83	1.99	0.13	0.86	1.99	-0.26	0.86
3	2.29	-0.08	0.82	1.89	0.12	0.88	1.82	-0.4	0.89
4	2.15	-0.02	0.84	2.0	0.20	0.86	1.94	-0.36	0.88
5	2.26	-0.03	0.82	1.72	0.06	0.90	1.72	-0.52	0.91

(b)

Model	LLR			BFGS (NN)			Back Propagation (NN)		
Combination	RMSE (m)	MBE (m)	R^2	RMSE (m)	MBE (m)	R^2	RMSE (m)	MBE (m)	R^2
1	2.06	0.60	0.89	2.34	0.72	0.84	2.21	-0.35	0.86
2	1.94	0.50	0.90	**1.91**	0.85	**0.90**	**1.83**	0.05	**0.88**
3	2.09	0.70	0.88	2.19	0.33	0.85	2.07	0.11	**0.88**
4	1.96	**0.47**	**0.91**	2.18	**0.21**	0.86	2.18	**-0.01**	0.85
5	**1.89**	0.72	0.91	3.04	1.35	0.76	2.31	-0.21	0.85

Bold values indicate best values.

made us confident about the selection of data length used for non linear modeling to counter the over fitting problems. Two different neural network methods, one trained with BFGS based quasi-Newton algorithm and the other based on Back Propagation algorithm were used for the development of prediction models. Developed Non linear models were also compared with the LLR based technique. Output results showed that LLR based techniques performed relatively better than ANN based techniques with relatively higher values of R^2 and lower values of RMSE and MBE for predicting both mean monthly reservoir levels. As far as ANN based techniques are concerned, Back Propagation based Neural Network model outperformed BFGS based radial ANN model with relatively good results.

For Local Linear Regression based modeling technique was found to be the best as compared to ANNs with the best input combination being the 4th one; $R^2 = 0.91$; (111110100111111; Table 2) having input combination (Inflow (T), Inflow (T-1), Inflow (T-2), Outflow (T), Outflow (T-1), Rainfall (T), Evaporation (T), Evaporation (T-1), Evaporation (T-2), Seepage (T), Seepage (T-1), Seepage (T-2)) for testing. Therefore, it can easily be said that along with Artificial Neural Network (ANN) based models, local linear regression LLR can also be used effectively for development of prediction models for effective water management. It is also recommended that similar studies should also be carried out on prediction of other hydro-meteorological variables so as to assess the ability of LLR based technique in modeling hydrological processes.

References

Agalbjorn, S., Koncar, N., and Jones, A. J. (1997). “A note on the gamma test.” *Neural Computing and Applications*, Vol. 5, No. 3, pp. 131-133, DOI: 10.1007/BF01413858.

Ahmadi, A., Han, D., Karamouz, M., and Remesan, R. (2009). “Input data selection for solar radiation estimation.” *Hydrol Processes*, Vol. 23, No. 19, pp. 2754-2764, DOI: 10.1002/hyp.7372.

Alqudah, A., Chandrasekar, V., and Le, M. (2013). “Investigating rainfall estimation from radar measurements using neural networks.” *Nat. Hazards Earth Syst. Sci.*, No. 13, pp. 535-544, DOI: 10.5194/nhess-13-535-2013.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). “Artificial neural networks in hydrology I: Preliminary concepts.” *J. Hydrol. Engng.*, ASCE, Vol. 5, No. 2, pp. 115-123, DOI: 10.1061/(ASCE) 1084-0699(2000).

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). “Artificial neural networks in hydrology II: Hydrologic applications.” *J. Hydrol. Engng.*, ASCE, Vol. 5, No. 2, pp. 124-137, DOI: 10.1061/(ASCE) 1084-0699(2000)5:2(124).

Chuzhanova, N. A., Jones, A. J., and Margetts, S. (1998). “Feature selection for genetic sequence classification.” *Bioinformatics*, Vol. 14, No. 2, pp. 139-143, DOI: 10.1093/bioinformatics/14.2.139.

Durrant, P. J. (2001). *winGamma: A non-linear data analysis and modelling tool with applications to flood prediction*, PhD Thesis, Department of Computer Science, Cardiff University, Wales, UK.

Evans, D. and Jones, A. J. (2002). “A proof of the gamma test.” *Proc. Roy. Soc. Series A*, Vol. 458, pp. 2759-2799, DOI: 10.1098/rspa.2002.1010.

Güldal, V. and Tongal, H. (2010). “Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in eğirdir lake level forecasting.” *Water Resource Management*, Vol. 24, No. 1, pp. 105-128, DOI: 10.1007/s11269-009-9439-9.

Han, D., Chan, L., and Zhu, N. (2007). “Flood forecasting using Support Vector Machines.” *Journal of Hydroinformatics*, Vol. 9, No. 4, pp. 267-276, 10.2166/hydro.2007.027.

Han, D., Cluckie, I. D., Karbassioun, D., Lawry, J., and Krauskopf, B. (2002). “River flow modelling using fuzzy decision trees.” *Water Resources Management*, Vol. 16, No. 6, pp. 431-445, DOI: 10.1023/A:1022251422280.

Hasebe, M. and Nagayama, Y. (2002). “Reservoir operation using the neural network and fuzzy systems for dam control and operation support." *Advances in Engineering Software*, Vol. 33, No. 5, pp. 245-260, DOI: 10.1016/S0965-9978(02)00015-7.

Hassan, M., Shamim, M. A., Hashmi, H. N., Ashiq, S. Z., Ahmed, I., Pasha, G. A., Naeem, U. A., Ghumman, A. R., and Han, D. (2014). "Predicting streamflows to a multipurpose reservoir using artificial neural networks and regression techniques." *Earth Science Informatics (In Press),* DOI: 10.1007/s12145-014-0161-7.

Ishak, A. M., Remesan, R., Srivastava, P. K., Islam, T., and Han, D. (2013). "Error correction modelling of wind speed through hydro-meteorological parameters and mesoscale model: A Hybrid approach." *Water Resources Management*, Vol. 27, No. 1, pp. 1-23, DOI: 10.1007/s11269-012-0130-1.

Jain, S. K., Das, A., and Srivastava, D. K. (1999). "Application of ANN for reservoir inflow prediction and operation." *Journal of Water Resource Planning and Management*, Vol. 125, No. 5, pp. 263-271, DOI: 10.1061/(ASCE)0733-9496(1999)125:5(263).

Jajarmizadeh, M., Lafdani, E. K., Harun, S., and Ahmadi, S. (2014). "Application of SVM and SWAT models for monthly Streamflow prediction, a case study in south of Iran." *KSCE Journal of Civil Engineering* (in press), DOI: 10.1007/s12205-014-0060-y.

Jones, A. J. (2004). "New tools in non-linear modelling and prediction." *Computational Management*, Science 1 pp. 109-149, 10.1007/s10287-003-0006-1.

Jones, A. J., Tsui, A., and de Oliveira, A. G. (2002). "Neural models of arbitrary chaotic systems: Construction and the role of time delayed feedback in control and synchronization." *Complexity International*, Vol. 9, p. tsui01: 1-9. Paper ID: tsui01, URL: http://www.complexity.org.au/vol09/tsui01/.

Kemp, S. E. (2006). *Gamma test analysis tools for non-linear time series*, PhD Thesis, Department of Computing and Mathematical Sciences University of Glamorgan UK.

Kisi, O. and Cimen, M. (2012). "Precipitation forecasting by using wavelet-support vector machine conjunction model." *Eng. Appl. of Artificial Intelligence*, Vol. 25, No. 4, pp. 783-792, DOI: 10.1016/j.engappai.2011.11.003.

Konèar, N. (1997). *Optimisation methodologies for direct inverse neurocontrol*, PhD Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London.

Lallahem, S. and Mania, J. (2003). "Evaluation and forecasting of daily groundwater outflow in a small chalky watershed." *Hydrological Processes*, Vol. 17, No. 8, pp. 1561-1577, DOI: 10.1002/hyp.1199.

Matheussen, B. V. and Thorolfsson, S. T. (1999). "Simulation errors due to insucient temporal resolution in urban snowmelt models." *in I. Jolie & J. Ball, eds, 'The 8th International Conference on Urban Storm Drainage (8ICUSD)'*, Vol. 4, pp. 2107-2114, 30 August - 3 September, Sydney, Australia.

McCulloch, W. S. and Pitts, W. (1943). "A logical calculus of the ideas imminent in nervous activity." *Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133, DOI: 10.1007/BF02478259.

Minsky, M. and Papert, S. (1969). "Perceptrons.*" MIT Press: Cambridg*e, MA.

Penrose, R. (1955). "A generalized inverse for matrices." *Proceedings of the Cambridge Philosophical Society*, Vol. 51, pp. 406-413, DOI: 10.1017/s0305004100030401.

Penrose, R. (1956). "On best approximate solution of linear matrix equations." *Proceedings of the Cambridge Philosophical Society*, Vol. 52, pp. 17-19, DOI: 10.1017/S0305004100030929.

Raman, H. and Chandramouli, V. (1996). "Deriving a general Operating Policy for reservoir using Neural Network." *Journal of Water Resources Planning and Management*, ASCE, Vol. 122, No. 5, pp. 342-347, DOI: 10.1061/(ASCE)0733-9496(1996)122:5(342).

Remesan, R., Shamim, M. A., Ahmadi, A., and Han, D. (2010). "Effect of data time interval on real-time flood forecasting." *Journal of Hydroinformatics*, Vol. 12, No. 4, pp. 396-407, DOI: 10.2166/hydro.2010.063.

Remesan, R., Shamim, M. A., and Han, D. (2008a). "Model data selection using gamma test for daily solar radiation estimation." *Hydrological Processes*, Vol. 22, No. 21, pp. 4301-4309, DOI: 10.1002/hyp.7044.

Remesan, R., Shamim, M. A., Han, D., and Mathew, J. (2008b). "ANFIS and NNARX based rainfall runoff modelling." *IEEE International Conference on Systems*, Man and Cybernetics, Singapore

Remesan, R., Shamim, M. A., Han, D., and Mathew, J. (2009). "Runoff prediction using an integrated hybrid modelling scheme." *Journal of Hydrology*, Vol. 372, No. 6, pp. 48-60, DOI: 10.1016/j.jhydrol.2009.03.034.

Rogers, L. L. and Dowla, F. U. (1994). "Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling." *Water Resources Research*, Vol. 30, No. 2, pp. 457-481, DOI: 10.1029/93WR01494.

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanics*, Spartan

Shamim, M. A., Bray, M., Remesan, R., and Han, D. (2014). "A hybrid modeling approach for assessing solar radiation." *Theoretical and Applied Climatology* (in press) DOI: 10.1007/s00704-014-1301-1.

Shamim, M. A., Ghumman, A. R., and Ghani, U. (2004). "Forecasting groundwater contamination using artificial neural networks." *International Conference on Water Resources and Arid Environments*, Riyadh Saudi Arabia.

Shamim, M. A., Remesan, R., Han, D., and Ghumman, A. R. (2010). "Solar radiation estimation in ungauged catchments." *Proceedings of ICE-Water Management*, Vol. 163, No. 7, pp. 349-359, DOI: 10.1680/wama.2010.163.7.349.

Srivastava, P. K., Han, D., Rico-Ramiraz, M. A., Bray, M., and Islam, T. (2012). "Selection of classification techniques for land use/land cover change investigation." *Advances in Space Research*, Vol. 50, No. 9, pp. 1250-1265, DOI: 10.1016/j.asr.2012.06.032.

Tayfur, G. and Guldal, V. (2006). "Artificial neural networks for estimating daily total suspended sediment in natural streams." *Hydrology Research*, Vol. 37, pp. 69-79, DOI: 10.2166/nh.2005.031.

Tokar, A. S. and Markus, M. (2000). "Precipitation-runoff modelling using artificial neural networks and conceptual models." *Journal of Hydrologic Engineering*, ASCE, Vol. 5, pp. 156-161, DOI: 10.1061/(ASCE)1084-0699(2000)5:2(156)).

Trigo, R. M. and Palutikof, J. P. (1999). "Simulation of daily temperatures for climate change scenarios over Portugal: A neural network approach." *Climate Research*, Vol. 13, pp. 45-59, DOI: 10.3354/cr013045.

Tsui, A. P. M. (1999). *Smooth data modelling and stimulus-response via stabilisation of neural chaos*, PhD Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London.

Tsui, A. P. M., Jones, A. J., and de Oliveira, A. G. (2002). "The construction of smooth models using irregular embeddings determined by a gamma test analysis." *Neural Computing and Applications*, Vol. 10, No. 4, pp. 318-329, DOI: 10.1007/s005210200004.

United Nations Edicational Scientific and Cultural Organisation (UNESCO) (2003), "World Water Development Report, 1st edition, Water for People, Water for Life."

Water and Power Development Authority, WAPDA (2010). "Kalabagh Dam Project report." pp. 12.