

# River Flow Forecasting using Neural Networks Coupled with Wavelet Analysis

Mukesh K. Tiwari

College of Agricultural Engineering and Technology  
Anand Agricultural University,  
Godhra – 389 001, India

**Abstract-** Daily river flow forecasting is an important component of effective and sustainable management of water resources. Accurate predictions of daily river flow can play a significant role for water resources planners and managers. Performance of traditional Neural Network models weakens with non-stationary dataset. To improve the NN model performance, a novel approach based on coupling discrete wavelet transforms (DWT) neural networks (NNs) for river flow forecasting is explored in this study. NNs-wavelet based NNs (WNNs), multiple linear regression (MLR) and wavelet based multiple linear regression (WMLR) models are developed in this study for river flow forecasting in the Upper Mahi river basin, Gujarat, India. The performance of the developed models is evaluated using the coefficient of determination, Nash-Sutcliffe coefficient, root mean square error, and mean average error. The key variables used to develop and validate the models are daily precipitation, daily maximum temperature and daily river flow. It is found in this study that the WNNs models are found to provide more accurate river flow forecasts than the NNs, WNNs, WMLR and MLR models. The results of this study indicate that coupled wavelet-higher order neural networks (WNNs) models improve the performance significantly and can be used successfully for accurate and reliable river flow forecasting.

**Keywords:** Higher order neural networks; wavelet; forecasting; Mahi river; Gujarat

## I. INTRODUCTION

Daily runoff forecasting is important for water resources planning and management such as reservoir operation, flood forecasting, canal operation, designing soil and water conservation structures, etc. Several rainfall-runoff models based on either physical or mechanistic approach, conceptual approach or on a system theoretic approach have been developed and successfully applied for runoff forecasting. Physically distributed modeling explicitly accounts for the small scale physics of the system but has been criticized due to high data requirement at different time and scale that create very complex models, which leads to the problems of over-parameterization and equifinality that can further increase forecast uncertainty. The main concern with system theoretic approach is that it does not consider system operation and the inherent physical processes. Neural networks (NNs), one of the system theoretic approaches has received considerable

attention for runoff forecasting in the last few decades. NN models are applied widely due to their capability to map complex nonlinear rainfall-runoff relationships, that is reflected by several successful applications in water resources.

All of the above cited studies used the general multi-layer feed-forward neural network (MLFF-NN) model coupled with the back-propagation error algorithm. These MLFF-NN models are first-order neural network (linear synaptic neural, LSN) and are capable to extract/capture only first-order correlations by employing a linear synaptic operation between the input vector and the synaptic weight vector (Giles and Maxwell, 1987, Gupta et al., 2003).

Besides excellent capacity to extract non-linearity from the input output mapping, NN models are criticized due to their limited ability to account for any physics of the hydrologic processes in a watershed. Daily runoff is widely perceived as non-linear and non-stationary. Non-stationarity that is reflected in terms of trends and seasonal variations influences the rainfall runoff transformation greatly and often results in poor predictability in operational applications. The physical processes associated with the rainfall runoff transformation greatly affect the streamflow generation for different periods as well. For instance, the low flows are generally associated by the base flows, whereas high flows are related with intensive rainfalls. In earlier studies it has been advocated that in the condition when non-stationarity limits the use of NN models, pre-processing of the input and/or output data can improve the NN model performance.

Wavelet transformation, that provides a time-frequency representation of a signal can give detailed information about the inherent physical structure of the data (Daubechies 1990). In the wavelet transformation technique the original signal is decomposed into sub-signal or sub-time series at different time and frequency, and these wavelet-transformed data provide different information at various resolution levels. Due to these capabilities of wavelet analysis it is widely applied to time series analysis of non-stationary signals. Wavelet transformation methodology has been used successfully in different discipline for analyzing variations, periodicities and trends in non-stationary signals (Xingang et al. 2003; Yueqing et al. 2004; Partal & Kucuk 2006). Wavelet analysis has also been applied in some of the water resources studies. There are several studies that uses capabilities of both neural networks and wavelet analysis to improve the model performance for water resource variables modeling and forecasting (Adamowski, 2008a, b; Satyajirao & Krishna, 2009; Wang & Meng, 2007; Tiwari and

Adamowski, 2013, Tiwari and Makwana, 2015, Kumar et al., 2015). In the present study the capabilities of ANN coupled wavelet analysis is explored for river flow forecasting in the Upper Mahi river basin, Gujarat, India.

*Wavelet Analysis*

The wavelet analysis decomposes the original time series data into a set of basis functions  $\{\psi_{a,b}(t)\}$  by translating and scaling the mother wavelet function  $\psi(t)$  that is mathematically represented as.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a > 0, \quad -\infty < b < \infty, \quad (1)$$

where  $a$  is the scale parameter, and  $b$  is the location parameter.. The mother wavelet  $\psi(t)$  is defined as (i)

$\int_{-\infty}^{+\infty} \psi(t)dt = 0$ , and (ii)  $\int_{-\infty}^{+\infty} \psi^2(t)dt = 1$ , satisfying that the function should have zero mean and be localized in both the time and frequency space. For a time series or finite energy signal  $f(t)$ , the continuous wavelet transform (CWT) is defined as

$$W_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} |\psi|\left(\frac{t-b}{a}\right) f(t)dt, \quad (2)$$

where  $|\psi|$  is the mother wavelet complex conjugate.

The discrete wavelet transform (DWT) is generally preferred in hydro-meteorological time series decomposition as these time series data are usually recorded in discrete time intervals. The DWT is obtained using dyadic sampling of  $W_{\psi}(a,b)$ , where the mother wavelet is scaled by powers of two viz.  $a = 2^j$  and translated by  $b = k2^j$ , where  $k$  is a location index and  $j$  is the decomposition level. In this way the DWT of  $f(t)$  is expressed as

$$T_{j,k} = 2^{-j/2} \sum_{i=0}^{N-1} f(t) \psi(2^{-j}t - k) \quad (3)$$

Where is the discreet wavelet coefficient;  $N$  is the length of data series which is an integer power of 2, i.e.,  $N=2^M$ . This gives the ranges of  $j$  and  $k$  as  $0 < k < 2^{M-j} - 1$  and  $1 < j < M$ , respectively. It shows that that at the largest scale (i.e.,  $2^j$  where  $j=M$ ), only one wavelet can cover the entire time interval generating a single coefficient. At the next scale ( $2^{j-1}$ ), two wavelets would cover the time interval producing two coefficients, and so on till  $j=1$ . Thus, the total number of coefficients generated by DWT for a discrete time series of length  $N = 2^M$  is  $1+2+3+\dots+2^{j-1} = N-1$  (Nourani et al., 2009).

The process consists of a number of successive filtering steps in which the time series is decomposed into approximation (A) and detail sub-time series or wavelet components (D1, D2, D3, etc). Approximation component represents the slowly changing coarse features of a time series and are

obtained by correlating stretched version (low-frequency and high-scale) of a wavelet with the original time series, while detail components signify rapidly changing features of the time series and are obtained by correlating compressed wavelet (high-frequency and low-scale) with the original time series.

*Study Area and Data Applied*

The present study was carried out at Limkheda agricultural watershed located in the semi-arid middle region of Gujarat, India (Fig. 1). The total area of the Limkheda watershed is 220.86 km<sup>2</sup>. The outlet of the study area is located at latitude 22° 49' 55" and longitude 73° 59' 15", falling within Survey of India toposheet Nos. F43I1, F43I2, F43H13 on 1:50,000 scale. The study area attains maximum elevation of 490 m and a minimum of 196 m above mean sea level. The climatic patterns in the watershed are characterized by wet summer and dry winter seasons, and very high temperatures throughout the year. The average depth of annual rainfall in the study area (Limkheda watershed) is 660 mm. As the watershed being situated in semi arid region and dominated with agriculture and forest land, water availability in the region is an important and critical issue. One rainwater harvesting structure (Umaria reservoir) has been put in place over the past years, having a bit success in improving the water availability for food production. There is huge scope of improving the potential of the watershed for increasing the availability of water for agriculture.

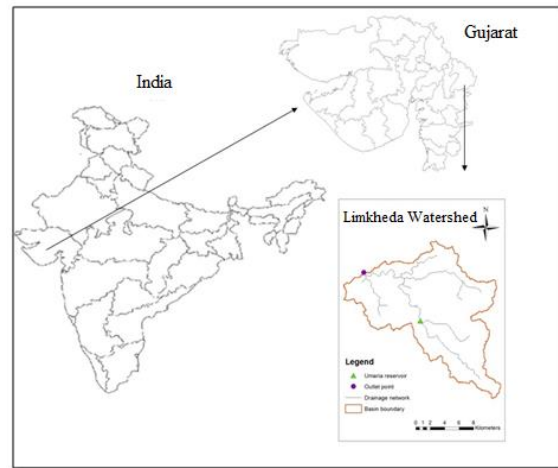


Fig. 1 Location map of Limkheda Watershed

In this study runoff data at the outlet of the watershed and release runoff data from the Umaria dam located upstream were collected during the monsoon period (1 June to 30 September) for 6 years from 2007 to 2012, and were selected for model development. For all the model development five years of data (2007-2011) were applied for the model training whereas one year data (2012) were applied for the evaluation of the developed models. Some of the statistical properties of these data are shown below.

Table 1: Some of the statistical properties of training and validation dataset

Length of Data	Data Patterns	Average (m <sup>3</sup> /s)	Min. (m <sup>3</sup> /s)	Max. (m <sup>3</sup> /s)	Std. (m <sup>3</sup> /s)	Skewness	Kurtosis
Training (2007-2011)	610	10.38	0.00	196.27	18.51	4.16	31.40
Validation (2012)	122	17.43	0.00	728.06	76.61	7.62	64.95

II. METHODOLOGY

Development of NN models

For development of an NN model a three layered feed forward back propagation neural network (FFBP-NN) was considered in this study. Selection of appropriate input variables is one of the important steps in NN model development. Considering that different models may have their own ability to map non-linear relationship between input variables and target variable, runoff data at the outlet of the watershed and from the Umaria dam with 1-3 days lags were considered to develop and evaluate different models as represented mathematically below:

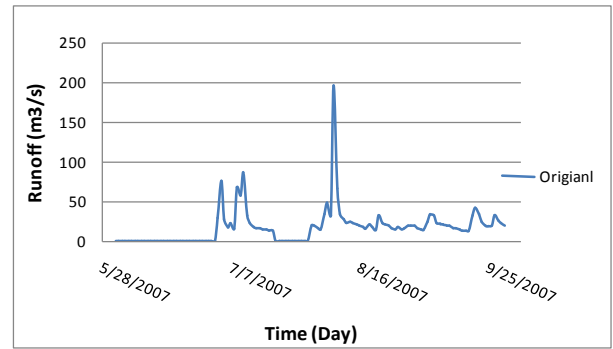
$$QL_{t+1} = f(QL_t, QL_{t-1}, QL_{t-2}; QU_t, QU_{t-1}, QU_{t-2}) \quad (4)$$

where, *QL* represents runoff at the outlet of the watershed, *QU* represents the runoff from the Umaria dam, whereas *t* represents the time.

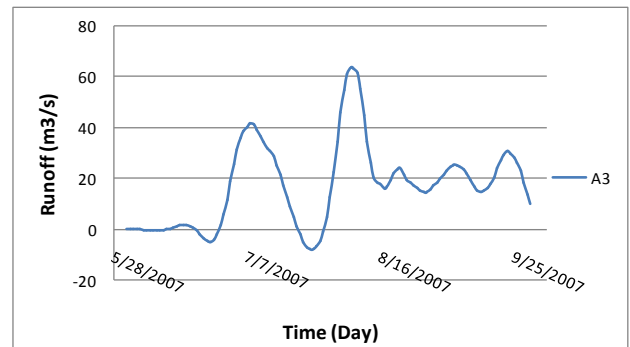
After considering these input variables, in the next step to select optimum number of hidden neurons, a trial and error procedure was used to ensure optimum NN model architecture. Levenberg-Marquardt (LM) training algorithm was used to achieve optimum values of weights.

Development of WNN models

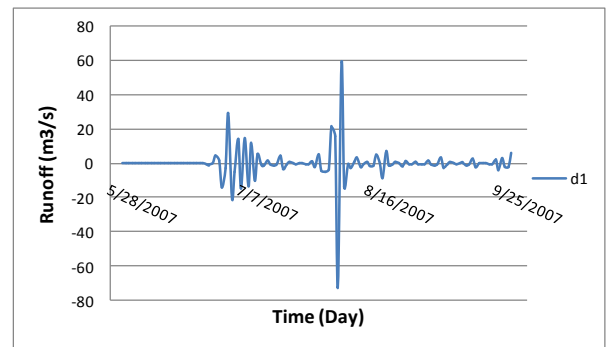
Runoff time series data at the outlet as well as at the outlet of the Umaria dam were decomposed using the DWT. The most widely tested and applied 'db5' mother wavelet from the Daubechies family was applied along with three level of decompositions viz. d1, d2 and d3 representing the details and another component A3 representing the approximation of the time series data. All the time series data from year 2007-2012 viz. runoff at the outlet of the watershed and runoff from the upstream Umaria project were decomposed using DWT, but for illustration purpose only the different wavelet components of runoff at the outlet of the Limkheda watershed for the year 2007 are presented in Fig. 2.



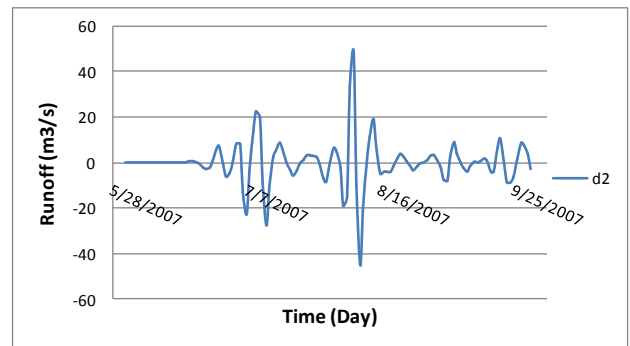
(a)



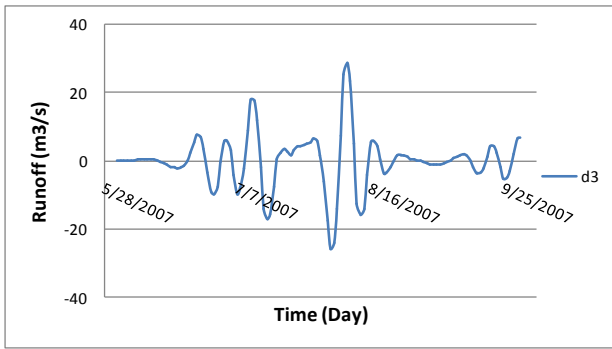
(b)



(c)



(d)



(e)

Fig. 2. Discrete wavelet components (a) Original (b) A3 (c) d1 (d) d2 and (e) d3 using DWT of runoff time series at the outlet of the Limkheda watershed

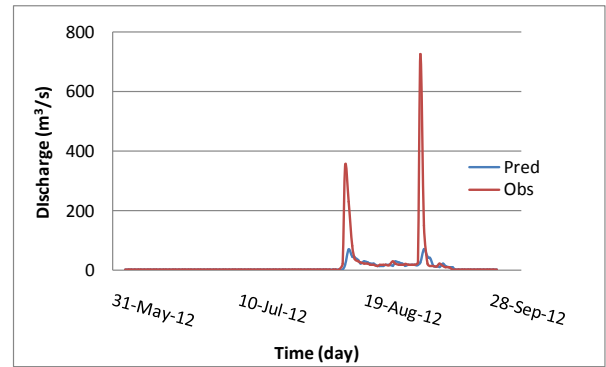
Even though the complexity of WNN models may be different and different input variables will play different role in both the NN and WNN models, but to benchmark the modelling capability of both the models similar input lagged variables viz. from 1 to 3 day were applied in WNN modelling, though the wavelet components of the respective variables were considered. Performance of both the models was also tested for 1-15 hidden neurons.

### III. RESULTS AND DISCUSSION

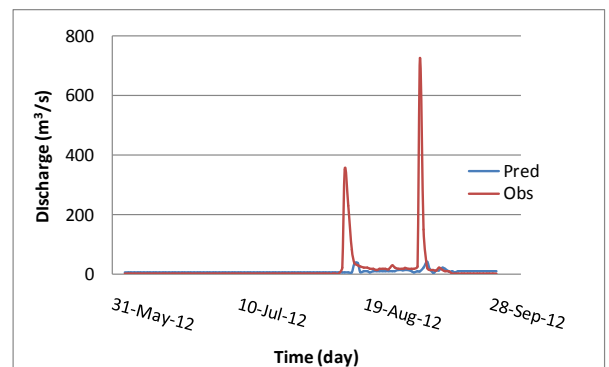
Performance of NN models is compared in terms of four performance indices as shown in Table 2. It can be observed from table that NN model performs better when runoff data from both the sources viz. runoff at the outlet and runoff from the Umariya dam are considered. In terms of different performance indices it can be observed that NN model is not able to perform outside the training dataset range and overall produce very poor performance indices. Better performance of NN model in terms of MAE compared to RMSE clearly indicates that it is able to simulate lower and medium runoff values but shows weakness in modelling extreme events. The poor performance of NN can also be observed from the observed and predicted values shown in Fig. 3.

Table 2: Performance of NN model during testing dataset

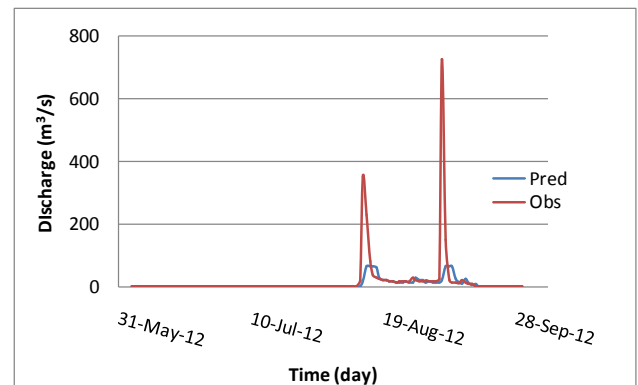
Model	Inputs	HN	E (%)	RMSE (m³/s)	Pdv (%)	MAE (m³/s)
1	QL(t-1;t-2;t-3)	2	9.65	73.38	90.51	13.67
2	QU(t-1;t-2;t-3)	1	-0.71	77.47	94.16	20.27
3	QL(t-1;t-2;t-3), QU(t-1;t-2;t-3)	1	8.42	73.88	90.92	14.93



(a) Model-1



(b) Model 2



(c) Model-3

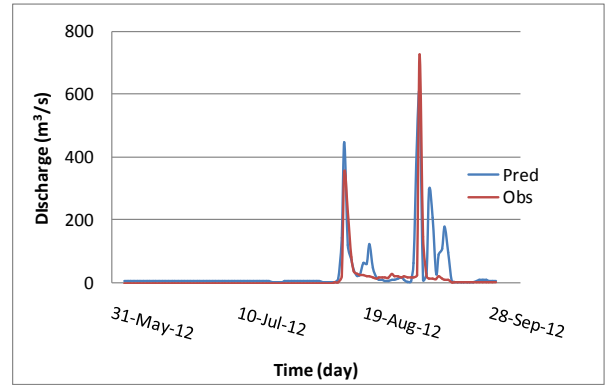
Fig.3. Performance of NN models for testing dataset using (a) Model 1, (b) Model 2, and (c) Model 3

All the time series data were decomposed using DWT and four wavelet sub time series viz. A3, d1, d2 and d3 were generated and used as input to the NN model to develop WNN models. Performance of these WNN models is presented below in Table 3. It can be observed from the table that WNN performs much better compared to NN model for runoff prediction. The best performance is obtained when all the wavelet components (Model #1) of runoff data at the outlet are considered. Such performance of WNN model becomes more important considering that validation dataset contains some extreme events. It further highlights that wavelet decomposition extract physical structure of the data and represents some of the physical processes associated with the runoff generation. Graph of observed and predicted

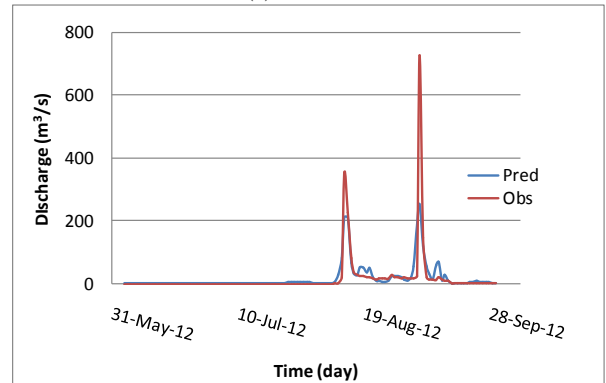
values during validation period is also shown in Fig. 4. It can be observed from the Figure that model #1 simulate the observed runoff values more precisely than the remaining models.

Table 3: Performance of WNN model

Model	Model inputs	Hidden Neurons	E (%)	RMSE (m <sup>3</sup> /s)	Pdv (%)	MAE (m <sup>3</sup> /s)
1	A3, d1, d2, and d3 of dis(t-1;t-2;t-3)	15	80.17	34.38	33.64	12.72
2	A3, d1, d2, and d3 of umdis(t-1;t-2;t-3)	2	0.57	76.98	89.77	21.13
3	A3, d1, d2, and d3 of dis(t-1;t-2;t-3), umdis(t-1;t-2;t-3)	14	36.38	61.58	17.91	22.91
4	A3, d1, d2, and d3 of dis(t-1;t-2)	3	59.13	49.36	65.17	13.58

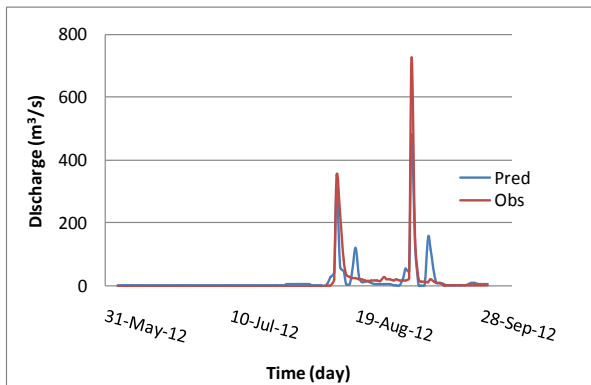


(c) Model 3

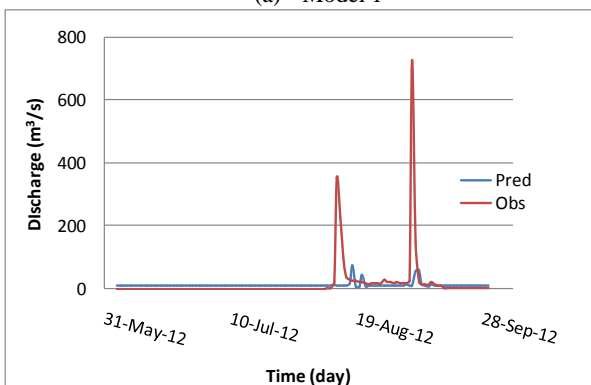


(d) Model 4

Fig.4. Performance of WNN models for testing dataset using (a) Model 1, (b) Model 2, (c) Model 3 and (d) Model 4



(a) Model 1



(b) Model 2

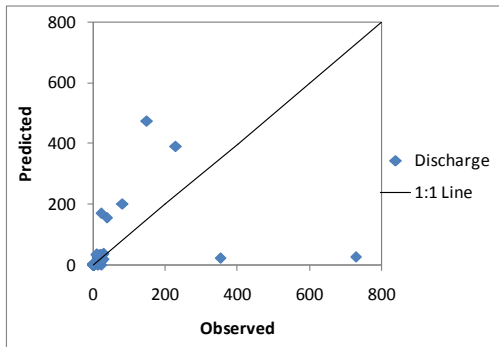
To benchmark the performance of previously discussed models, simpler models viz. MLR and WMLR were also developed using the same input variables as used for NN and WNN model development. The performance of these models is presented in terms of different performance indices in Table 3.

It can be observed that the performance of both NN and HONN models are better than MLR models, whereas WMLR and WHONN models perform very close to each other but their performance is slightly inferior compared to WNN model.

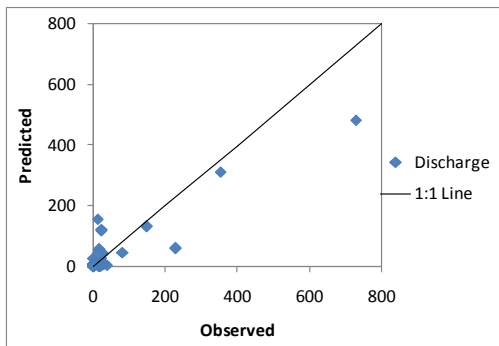
Table 4: Performance of MLR and WMLR models for testing dataset

Model	Input	E (%)	RMSE (m <sup>3</sup> /s)	Pdv (%)	MAE (m <sup>3</sup> /s)
MLR	dis(t-1;t-2;t-3)	-3.71	78.62	43.80	17.68
WMLR	A3, d1, d2, and d3 of dis(t-1;t-2;t-3)	75.39	38.30	31.56	15.21

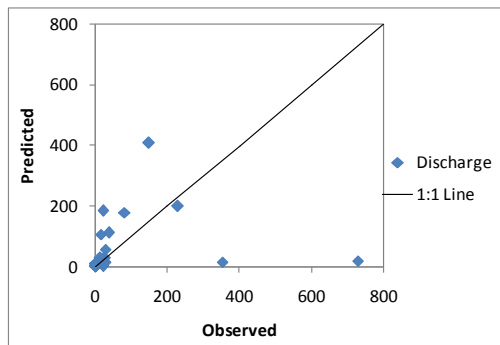
As the testing dataset contains a wide response of the watershed from 0 values to very extreme values, and therefore to further analyse the performance of all the models scatter plots are generated between observed and predicted runoff values as presented in Fig. 5



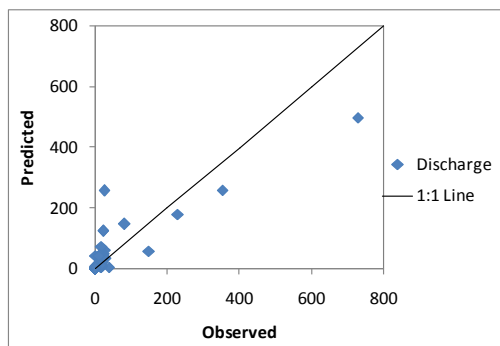
(a) best NN model



(b) best WNN model



(c) MLR model



(d) WMLR model

Fig. 5 Scatter plots of observed and predicted values using (a) best NN, (b) best WNN, (c) MLR, and (d) WMLR models.

#### IV. CONCLUSION

Performance of wavelet analysis based neural networks (WNNs) for riverflow forecasting is assessed in this study. In terms of different performance indices it is find in this study that WNN models provide more accurate river flow forecasts than the NNs, WNNs, WMLR and MLR models. It is observed in this study that wavelet based models such as WNN and WMLR models simulate the epak discharge values better than traditional NN and MLR models. Overall, results of this study indicate that WNNs models improve the performance significantly and can be used successfully for accurate and reliable river flow forecasting.

#### V. REFERENCES

- [1] Adamowski, J. F. 2008a River flow forecasting using wavelet and crosswavelet transform models. *Hydrol. Processes* 22(25), 4877–4891.
- [2] Adamowski, J. F. 2008b Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *J. Hydrol.* 353(3–4), 247–266.
- [3] Daubechies, I., 1990. The wavelet transform, time–frequency localization and signal analysis. *IEEE Transactions on Information Theory* 36 (5), 6–7.
- [4] Giles, L., and Maxwell, T. (1987). “Learning, invariance and generalization in high-order neural networks.” *Appl. Opt.*, 26(23), 4972–4978.
- [5] Gupta, M. M., Jin, L., and Homma, N. (2003). *Static and dynamic neural networks: From fundamentals to advanced theory*, Wiley, New York.
- [6] Kumar, S., Tiwari, M.K., Chatterjee, C., Mishra, A. (2015). Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method. *Water Resources Management*, DOI: 10.1007/s11269-015-1095-7. (Impact Factor: 2.600).
- [7] Makwana, J., Tiwari, M.K., (2015). Prioritization of agricultural sub-watersheds in semi arid middle region of Gujarat using Remote Sensing and GIS. *Environmental Earth Sciences*. (*Impact Factor: 1.572*).
- [8] Nourani, V., Komasi, M., and Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water Resources Management*, 23 (14), 2877–2894. doi:10.1007/s11269-009-9414-5.
- [9] Partal, T., Kucuk, M., 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Physics and Chemistry of the Earth* 31, 1189–1200.
- [10] Satyajirao, Y. R. & Krishna, B. 2009 Modelling hydrological time series data using wavelet neural network analysis. *IAHS Publ.*
- [11] Tiwari, M.K. and Adamowski, J.F., 2013. Urban water demand forecasting and uncertainty assessment using ensemble wavelet–bootstrap neural network models. *Water Resources Research*, 49 (10), 6486–6507.
- [12] Wang, J. & Meng, J. 2007 Research on runoff variations based on wavelet analysis and wavelet neural network model: a case study of the Heihe River drainage basin (1944–2005). *J. Geog. Sci.* 17(3), 327–338.
- [13] Xingang, D., Ping, W., Jifan, C., 2003. Multiscale characteristics of the rainy season rainfall and interdecadal decaying of summer monsoon in North China. *Chinese Science Bulletin* 48, 2730–2734.
- [14] Yueqing, X., Shuangcheng, L., Yunlong, C., 2004. Wavelet analysis of rainfall variation in the Hebei Plain. *Science in China Series D Earth Science* 48, 2241–2250.