J. Indian Water Resour. Soc.,
Vol. 39, No. 2, April, 2019

FLOOD FORECASTING USING HYBRID WAVELET NEURAL NETWORK MODEL

R Venkataramana*[1], V. S. Jeyakanthan[1], Y. R. Satyaji Rao[1] and T. Vijay[1]

ABSTRACT

The dynamic and accurate flood forecasting of daily stream flow processes of a river are important in the management of extreme events such as flash floods, floods and optimal design of water storage structures and drainage network. This paper aims to recommend a best hydrologic models are linear stochastic models autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and nonlinear models like Artificial neural network (ANN) and Wavelet neural network (WNN) for flood forecasting of Vamsadhara river in the India (sub zone 4A). Wavelet neural network (WNN) is an hybrid modelling approach for forecasting of river flow using daily time series data. A discrete wavelet multi resolution method was employed to decompose the time series data of river flow into sub series with low (approximation) and high (details) frequency, and these sub series were then used as input data for the artificial neural network (ANN). WNN model with db3 wavelet decomposition level was employed to forecast flood ahead of time. Daily flow data was collected from India-WRIS and rainfall from IMD from 1971 to 2013. 60% data was used for model calibration and 40% for validation. The one day ahead forecasting performances of all models for predicted period were compared. The comparison of model forecasting performance was conducted based upon different statistical indices and graphical criteria. The result indicates that WNN model is better than ANN, ARIMA and ARMA.

Keywords: *Forecast, Flood, Network. Model, Wavelet.*

INTRODUCTION

Floods are considered as natural disasters that can cause casualties and destraction to infrastructures. The problem is that floods, particularly in hilly topography sub zone 4(a), takes place in a very short time. So, it is important to forecast floods events earlier with a lead time up to 48 hours to give early warning alert to avoid or minimize disasters. Time series prediction of a flood data or any other fields that falls under time series category in a real scenario is much complex than non time series prediction. Recently, various soft computing techniques like Fuzzy logic, Artificial Neural Networks (ANN) and Genetic programming (GP) etc. were used efficiently in time series prediction to improve the forecasting accuracy. Soft computing techniques normally utilizes tolerance to uncertainties, imprecision, and partial truth associated with input information in order to cope up the draw backs in mathematical models. Process-based models apply physical principles to model various constituents of physical processes of the hydrological cycle. The black-box data-driven models, on the other hand, are primarily based on the measured data and map the input-output relationship without giving consideration to the complex nature of the underlying process. Among data-driven models, the ANNs have appeared as powerful black-box models and received a great attention during last two decades. The merits and shortcomings of using ANNs in hydrology can be found in ASCE Task Committee (2000a, b) and Abrahart et al. (2012). Despite good performance of the ANNs in modelling of non-linear hydrological relationships, yet these models may not be able to cope with non-stationary data if pre-processing of input and/or output data is not performed (Cannas et al. 2006). Application of wavelet transform (WT) has been found effective in dealing with this issue of non-stationary data (Nason and Von Sachs 1999). The WT is a mathematical tool that improves the performance of hydrological models by simultaneously considering both the spectral and the temporal information contained in the data. It decomposes the main time series data into its sub-components. Thus, the hybrid wavelet data driven models which use multi scale-input data, result in improved performance by capturing useful information concealed within the main time series data. Recently, various hydrological studies successfully applied WT to increase forecasting efficiency of neural network models. Partal (2016), Wei et al. (2013), Ramana. et.al (2013) and Anctil and Tape (2004) applied wavelet based data driven models for stream flow forecasting. Potential of hybrid wavelet-ANN (wavelet-Artificial neural network) to model flood events in the Achankovil river basin in Kerala. Hourly water level at the Konni gauging site is predicted with 1-, 3- and 6-h lead times and the capability of the hybrid model is proved by comparison with a simple ANN model. The WANN models developed were able to predict the magnitude of peak value and the time to peak value more accurately when compared to the ANN models (Alexznder. et, al 2018). A novel wavelet-artificial neural network hybrid model (WA-ANN) for short-term daily inflow forecasting is being adopted for the first time using Tropical Rainfall Measuring Mission (TRMM) data together with inflow data and transformed using mother-wavelets to improve the model performance (Celso,. et.al, 2019).

STUDY AREA

Vamsadhara River Basin is prone to frequent floods with an aerial extent of 10,601.5 sq.kms. This is an interstate drainage basin between Andhra Pradesh and Orissa. The Vamsadhara River basin witnesses a tropical hot climate which is influenced by the southwest monsoons during the months of June through October. Occasionally, cyclones strike the basin due to the formation of depression in the Bay of Bengal. The average depth of rainfall is 1302. About

1. *Scientist, NIH, DRC, Kakinada, AP-533 003,India,*
E-mail:venkataramana_1973@yahoo.co.in
Manuscript No. 1504

84.6% of rainfall was received during the five months monsoon period (June to October).The study area is located between longitudes 83^0 15' and 84^0 57' E and between latitudes 18^0 15' and 19^0 57' N (Figure. 1).

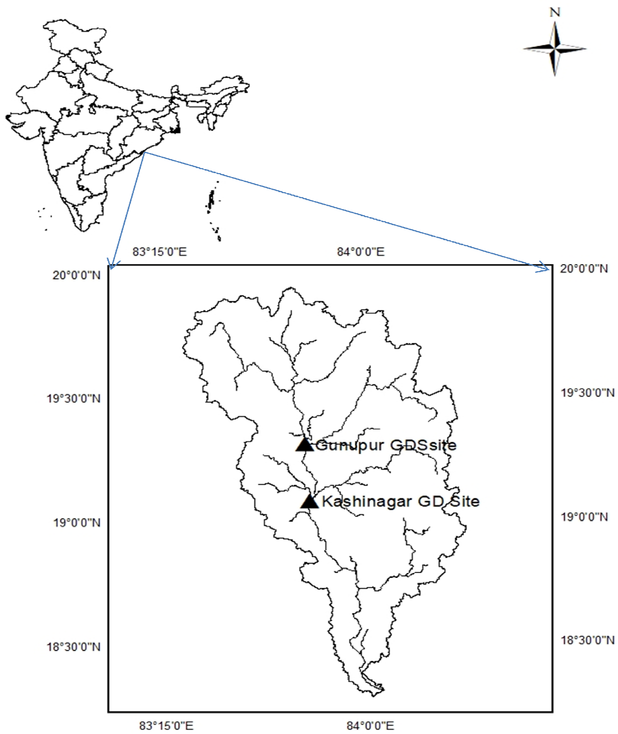


Fig. 1: Study Area of Vamasadhara River with GD Locations

METHODOLOGY

Box and Jenkins developed ARIMA stochastic models that describe a wide class of models forecasting a univariate time series that can be made stationary by applying transformations – mainly differences for Trend and Seasonality, and power function to regulate the variance (Box and Jenkins, 1970; Box and Jenkins, 1976) The model, word "ARIMA" consists of three terms i.e. i) AR ii) I and iii) MA terms. Lags of differenced time series in the forecasting equations are called "autoregressive(AR)" term, whereas lags of the forecasted errors are called "moving average (MA)" term and the time series which requires differenceation to become stationary should be "Integrated (I)" (Ghafoor and Hanif, 2005). AR(p), MA(q) and Auto regressive moving average (ARMA(p,q)) models are some special cases of Box and Jenkins ARIMA model. In this study, ARMA and ARIMA models are used to evaluate the performances of time series of rainfall and stream flow.

Artificial neural network: ANNs modeling is a branch of artificial intelligence developed in 1950s aiming at imitating the biological brain architecture. ANNs are parallel-distributed systems made of many interconnected nonlinear processing elements (PEs), called neurons (Hect-Nielsen, 1990). A renewal of interest has grown exponentially in the last decade, mainly concerning the availability of suitable hardware that has made them convenient for fast data analysis and information processing. Multi Layer Perceptron model (MLP) is the most commonly used type of ANNs. Its structure consists of PEs and connections. The PEs called neurons are arranged in layers: the input layer, one or more hidden layers, and the output layer. An input layer serves as buffer that distributes input signals to the next layer, which is a hidden layer. Each unit in the hidden layer sums its input, processes it with a transfer function, and distributes the result to the output layer. Also, several hidden layers are possibly connected in the same fashion. The units in the output layer operate in a similar manner. Such kind of models are defined as feed-forward ANNs, since data flows within the network, from one layer to the next without any return path.

Wavelet Transform: The properties of irregularity in shape and compactness make wavelets an ideal tool for analysis of non stationary signals. Fourier analysis decomposes a signal into sine and cosine waves of various frequencies whereas wavelet analysis decomposes a signal into shifted and scaled versions of the mother wavelet. The shifting (delaying) of the mother wavelet provides local information of the signal in time domain whereas scaling (stretching or compressing) of the mother wavelet provides local information of the signal in frequency domain (Daubechies, 1992). The scaling and shifting operations applied to mother wavelet are used to calculate wavelet coefficients that provide correlation between the wavelet and local portion of the signal. From the calculated wavelet coefficients, we can extract two types of components: approximate coefficients and detail coefficients. The approximate coefficients represent high scale, low frequency component of the original signal whereas detail coefficients represent low scale, high frequency component. Continuous wavelet transform (CWT) operates at every scale from that of the original signal up to some maximum scale. This distinguishes CWT from DWT (which operates at dyadic scales only). CWT is also continuous in terms of shifting: during computation, the analyzing wavelet is shifted smoothly over the full domain of signal. The results of the CWT are wavelet coefficients, which are a function of scale and position. Multiplying each coefficient by the appropriately scaled and shifted wavelet gives the constituent wavelets of the original signal. The computation of wavelet coefficients at every scale requires large computational time. To reduce the time, it is preferred to calculate wavelet coefficients for selected subset of scales and positions. If the scales and positions are selected based on power of two (dyadic scales and positions), then the analysis will be efficient and just as accurate, named discrete wavelet transform(DWT) (Mallat, 2009). The process of decomposition can be iterated, with successive approximations being decomposed in turn (discarding detail coefficients), so that original signal is broken down into many lower-resolution components. This process is referred as multi resolution analysis. We have selected Db3 (length-4 Daubechies) (Daubechies, 1992) wavelet as mother wavelet because this is one of the commonly used wavelets for separating fluctuations from the given time series. The smoothness of different wavelets depends on the number of vanishing moments (Mallat, 2009). Db3 wavelet has four vanishing moments, a smallest length wavelet with smoothness property. We have set maximum resolution level to value 10 for decomposition of every meteorological time series. The forward discrete wavelet transform is employed to decompose original time series of every meteorological variable at different scale (maximum level n = 10). The wavelet transform produces high-pass (detail) coefficients at every level and one low-pass (approximation) coefficient. The inverse discrete wavelet transform is applied on the produced coefficients at every level to generate DW subseries, which are of the same length as the original series. Thus, we obtain $n + 1$ (n detail and one approximate) DW subseries for the original time series of every meteorological variable. The correlation coefficient between the generated DW subseries at different level and the original rainfall series is calculated. The number of DW subseries which have high correlation with the original rainfall series is identified and summed up to generate a new (final) subseries for that meteorological variable. The objective behind addition of DW subseries having high correlation with the original rainfall and stream flow series is to reduce the number of variables (dimensions or inputs) and to increase the correlation between newly generated subseries and the original rainfall and stream flow series. This process is repeated for every rainfall and stream flow time series.

Selection of mother wavelet function: The performance of different wavelet coupled based models is very sensitive to the selection of the mother wavelet function to be used for transformation of data using DWT. A number of wavelet families are available, each having different members. Examples of these families include the most popular Daubechies db wavelet family containing db2, db3, db4, db5, db6, db7, db8, db9 and db10 members and the Coiflet wavelet family having Coif1, Coif2, Coif3, Coif4, Coif 5 as members (Daubechies and Bates 1993). These different wavelet functions are characterized by their distinctive features including the region of support and the number of vanishing moments. The region of support of a wavelet is associated with the span length of the wavelet which affects its feature localization properties of a signal and the vanishing moment limits the wavelet's ability to represent polynomial behavior or information in a signal. The details of different wavelet families can be found in many text books such as Daubechies and Bates (1993) and Addison (2002). The present study employed the db3 wavelet function to decompose the input rainfall and stream flow data. The db3 wavelet function with eight vanishing moments has the ability to best describe the temporal and the spectral information in the input rainfall and stream flow data (Shoaib et al. 2014, 2015, 2016).

RESULTS AND DISCUSSION

As per CWC report three basin are falls in flood prone area in the sub zone 4(A) of the Andhra region namely Vamsadhara river basin. Daily gauge discharge data downloaded from India-WRIS and IMD 0.5 degree gridded daily rainfall of Vamsadhara basin gauge discharge stations Kashinagar from 1971-2013 and Gunupu from 2003–2013 has been used in the analysis. 60% of the data was used for calibration and 40%of the data for validation for each gauge discharge (GD) station in the model. The performance of various models during calibration and validation were evaluated by using the statistical indices namely: the Root Mean Squared Error (RMSE), Correlation Coefficient (R) and Coefficient of Efficiency.

The definitions of different statistical indices are presented below:

$$\mathrm{RMSE} = \sqrt{\frac{\left[\sum_{t=1}^{N}(y_t^o - y_t^c)^2\right]}{N}} \tag{1}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{N}(y_t^o - \overline{y^o})(y_t^c - \overline{y^c})}{\sqrt{\left[\sum_{t=1}^{N}(y_t^o - \overline{y^o})^2\right]\left[\sum_{t=1}^{N}(y_t^c - \overline{y^c})^2\right]}} \tag{2}$$

$$\mathrm{COE} = 1 - \frac{\left[\sum_{t=1}^{N}(y_t^o - y_t^c)^2\right]}{\left[\sum_{t=1}^{N}(y_t^o - \overline{y^o})^2\right]} \tag{3}$$

where y_t^o and y_t^c are the observed and calculated values at time t respectively, $\overline{y^o}$ and $\overline{y^c}$ are the mean of the observed and calculated values.

The original time series was decomposed into Details and Approximations to certain number of sub-time series $\{D_1, D_2 \ldots D_p, A_p\}$ by wavelet transform algorithm. These play different role in the original time series and the behaviour of each sub-time series is distinct (Wang and Ding, 2003). So the contribution to original time series varies from each successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components, tested using different scales from 1 to 10 with different sliding window amplitudes. In this context, dealing with a very irregular signal shape, an irregular wavelet, the Daubechies wavelet of order 3 (DB3), has been used at level 3. Consequently, D_1, D_2, D_3 were detail time series, and A_3 was the approximation time series.

An ANN was constructed in which the sub-series $\{D_1, D_2, D_3, A_3\}$ at time t are input of ANN and the original time series at $t + T$ time are output of ANN, where T is the length of time to forecast. The input nodes for the WNN are the decomposed subsets of antecedent values of the rainfall and runoff are presented in Table 1. The Wavelet Neural Network model (WNN) was formed in which the weights are learned with Feed forward neural network with Back Propagation algorithm. The number of hidden neurons for BPNN was determined by trial and error procedure. The performance of various models estimated to forecast the river flow are presented in Table 2.

Table 1: Model Inputs for WNN

Model	Input Variables
I	Q(t) = f (Q [t-1])
II	Q(t) = f (Q [t-1], R [t-1])
III	Q(t) = f (Q [t-1], Q [t-2], R [t-1])
IV	Q(t) = f (Q [t-1], Q [t-2], R [t-1], R [t-2])
V	Q(t) = f (Q [t-1], Q [t-2], Q [t-3], R [t-2], R [t-1])
VI	Q(t) = f (Q [t-1], Q [t-2], Q [t-3], Q [t-4], R [t-3], R [t-2], R [t-1])

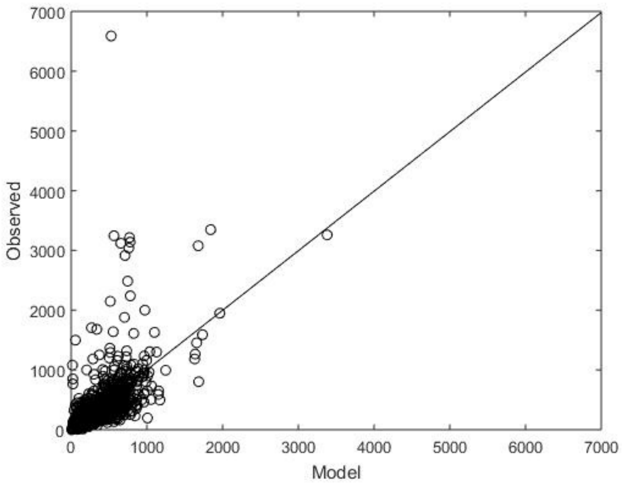
Note: Q is discharge and R is Rainfall.

From Table 2, it is found that low RMSE values of Vamsadhara basin GD sites of Kashinagar (87.6550 to 59.4632 cums) and Gunupur (59.5191 to 48.0763 cums) for WNN models when compared to ANN , ARIMA and ARMA models. It has been observed that WNN models estimated the peak values of discharges to a reasonable accuracy. Further, it is observed that the WNN model having three antecedent values of the time series, estimated minimum RMSE of Vamisidhara GD sites Kashinagar(0.9595Cums) and Gunupur(0.9272 cums) high correlation coefficient and highest percentage of efficiency (Kashinagar >91, Gunupur >84) during the validation period. Model IV and V for Vamisidhara GD sites of Kashinagar and Gunupur, Model of WNN was selected as the best fit model to forecast the river flow are day in advance.

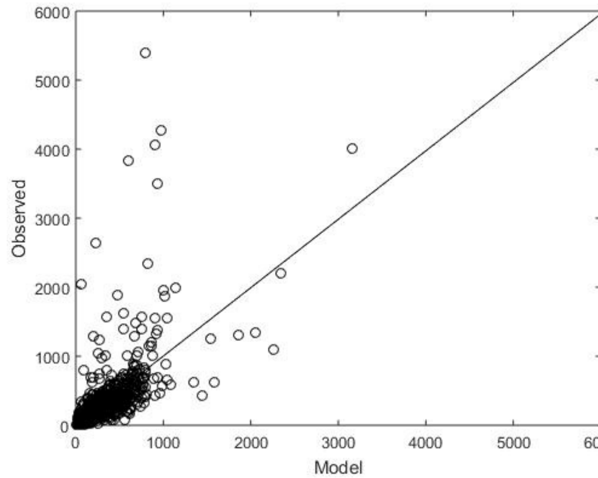
Table 2: Goodness of fit statistics of the calibration and validation the forecasted rainfall

Model		Calibration			Validation		
		RMSE	R	COE	RMSE	R	COE
Vamsadhara (Kashinagarr)							
ARMA							
		139.784	0.6975	47.6003	155.982	0.6483	40.4824
ARIMA							
		143.456	0.7682	48.8976	148.09	0.6783	42.3452
ANN							
Model	I	131.75	0.7312	53.4505	146.006	0.6922	47.8594
Model	II	125.519	0.7601	57.7493	140.525	0.7191	51.7075
Model	III	121.913	0.7755	60.1415	140.811	0.7214	51.5179
Model	IV	119.984	0.7836	61.3927	136.122	0.7398	54.6993
Model	V	119.176	0.7869	61.9107	131.773	0.7589	57.5539
Model	VI	117.553	0.7934	62.9407	142.946	0.7106	50.0577
WNN							
Model		76.7682	0.9179	84.1957	87.655	0.9063	81.2073
Model	II	59.6406	0.9513	90.461	88.2994	0.9148	80.9328
Model	III	44.262	0.9735	94.7461	61.1728	0.9574	90.8499
Model	**IV**	**43.0368**	**0.975**	**95.0329**	**59.4632**	**0.9595**	**91.3554**
Model	V	37.7801	0.981	96.1722	79.9264	0.929	84.3842
Model	VI	35.833	0.9827	96.5565	81.9237	0.9223	83.5963
Vamsadhara (Gunupur)							
ARMA							
		230.465	0.5774	31.4296	80.5468	0.8033	57.7296
ARIMA							
		229.567	0.5782	29.0893	84.5672	0.82134	59.4562
ANN							
Model	I	216.499	0.6287	39.5062	76.9465	0.803	61.3207
Model	II	166.039	0.8043	64.429	90.3716	0.7752	46.5047
Model	III	157.693	0.8244	67.9168	84.9136	0.7972	52.7449
Model	IV	142.432	0.8345	68.9562	86.7892	0.7862	54.6781
Model	V	149.176	0.7869	61.9107	90.6754	0.7589	57.5539
Model	VI	151.75	0.7312	53.4505	86.0059	0.6922	47.8594
WNN							
Model	I	103.645	0.9282	86.1356	59.5191	0.8781	76.8573
Model	II	70.1224	0.9679	93.6556	62.1344	0.8651	74.7119
Model	III	50.5531	0.9835	96.7028	57.4773	0.8861	78.3485
Model	IV	40.6531	0.9893	97.8677	51.0849	0.9131	82.8954
Model	**V**	**40.6419**	**0.9894**	**97.8688**	**48.0763**	**0.9272**	**84.8564**
Model	VI	32.3385	0.9933	98.6506	48.9779	0.9203	84.2914

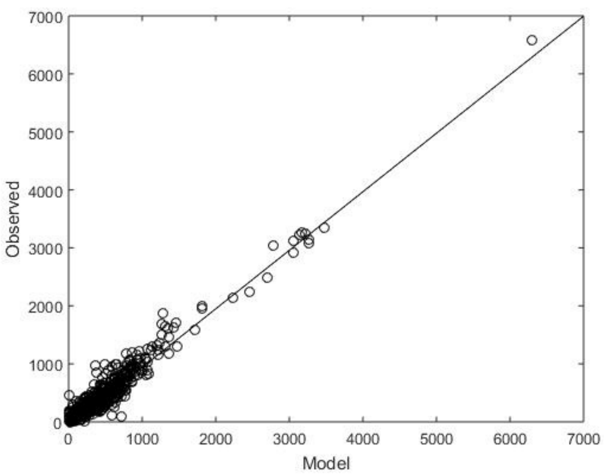
Figure 2 shows the scatter plot between the observed and modeled flows by WNN and ANN. It was observed that the flow forecasted by WNN models were very much close to the 45 degrees line. From this analysis, it was worth to mention that the performance of WNN was much better than ANN , ARIMA and ARMA models in forecasting the river flow in one-day advance.



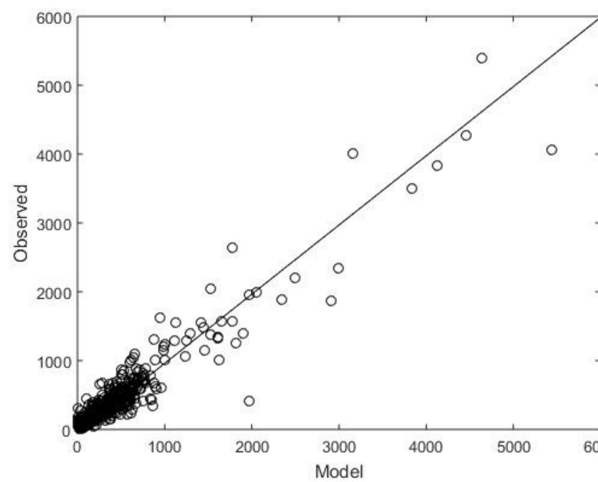
Kashinagar ANN calibration



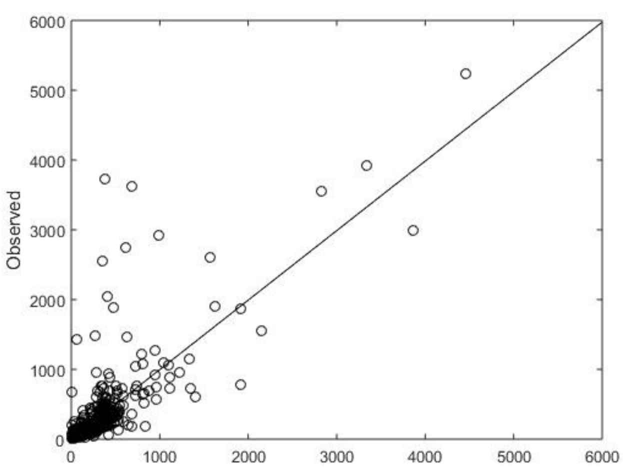
Kashinagar ANN validation



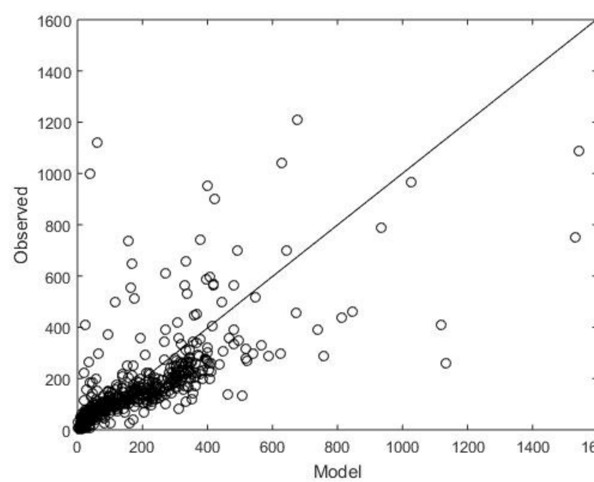
Kashinagar WNN calibration



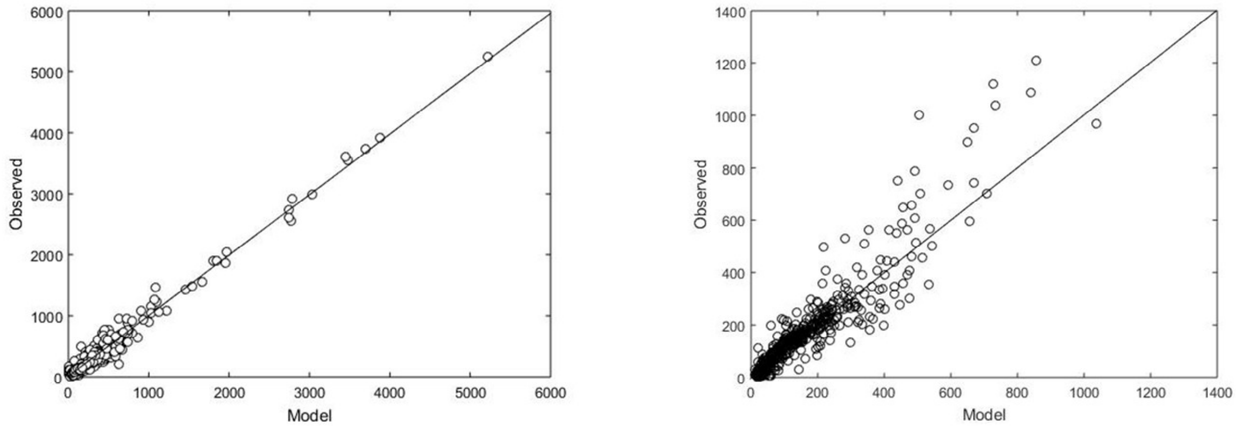
Kashinagar WNN validation



Gunupu ANN calibration



Gunupur ANN validation

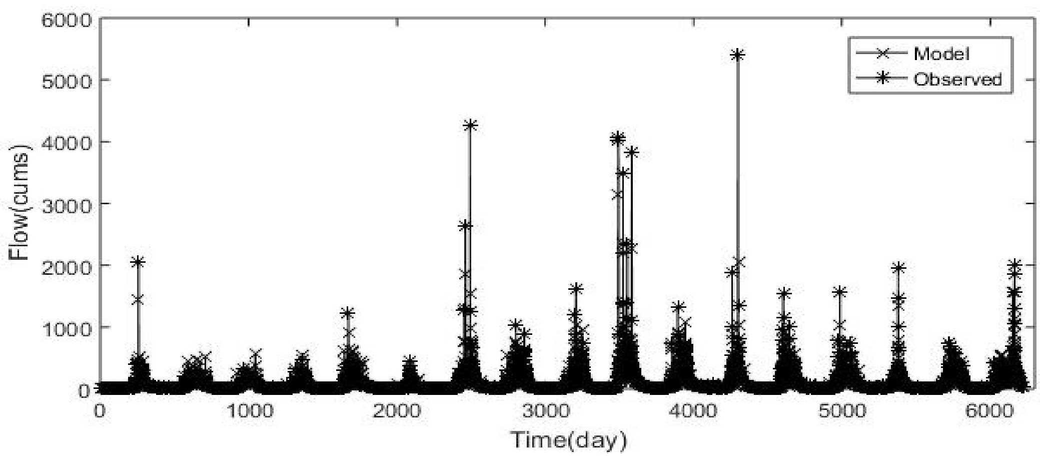


Gunupur WNN calibration **Gunupur WNN validation**

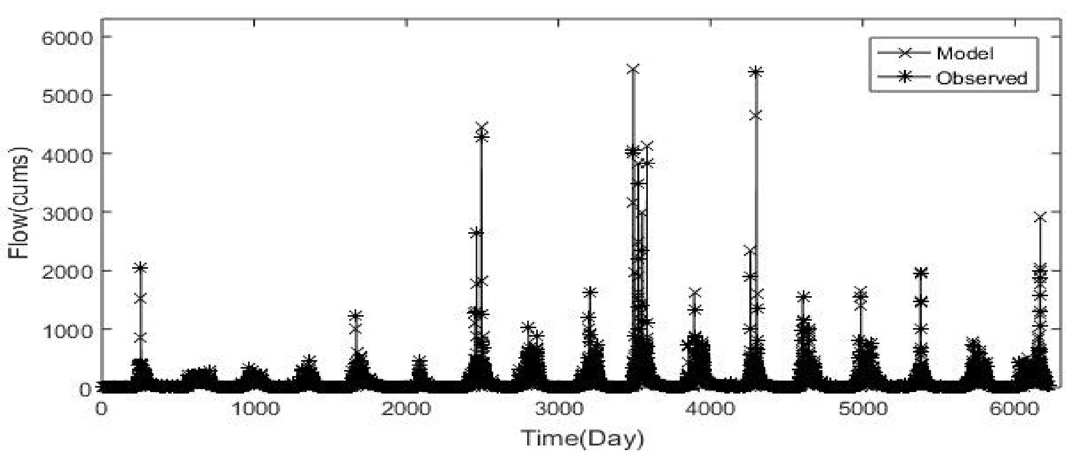
Fig. 2: Shows the scatter plot between the observed and modeled flows by WNN and ANN.

Figures 3 and 4, shows the observed and model graphs for ANN and WNN models during validation respectively. It is found that model values from WNN model properly matched with the observed values, whereas ANN model underestimated the observed values. From this analysis, it is evident that the performance of WNN was much better than ANN and AR models in forecasting the flood.

The distribution of error along the magnitude of rainfall computed by WNN and ANN models during the validation period has been presented in Figure. 5. From figure. 5, it was observed that the estimation of peak flood was very good as the error is minimum when compared with ANN model

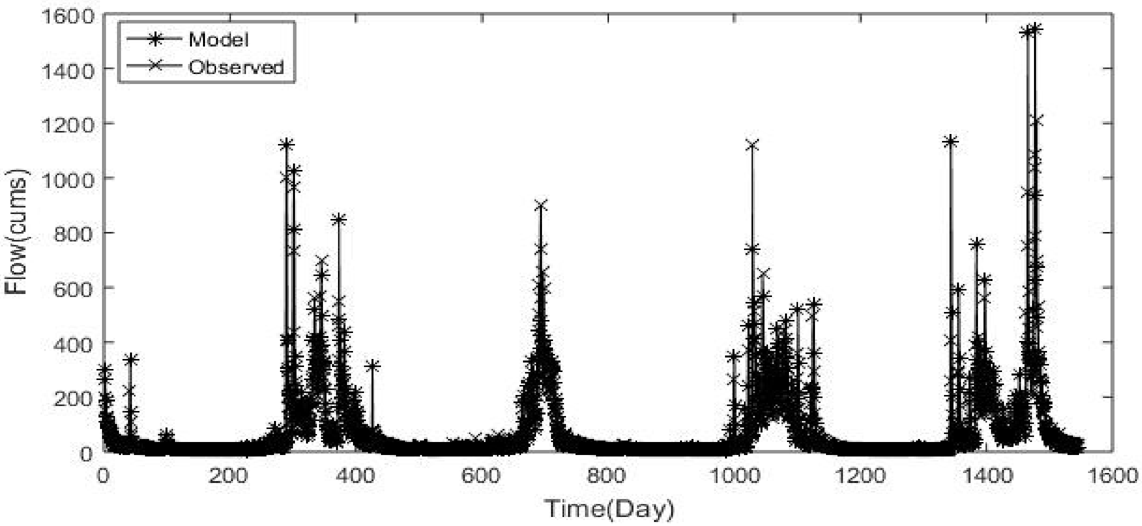


(a)

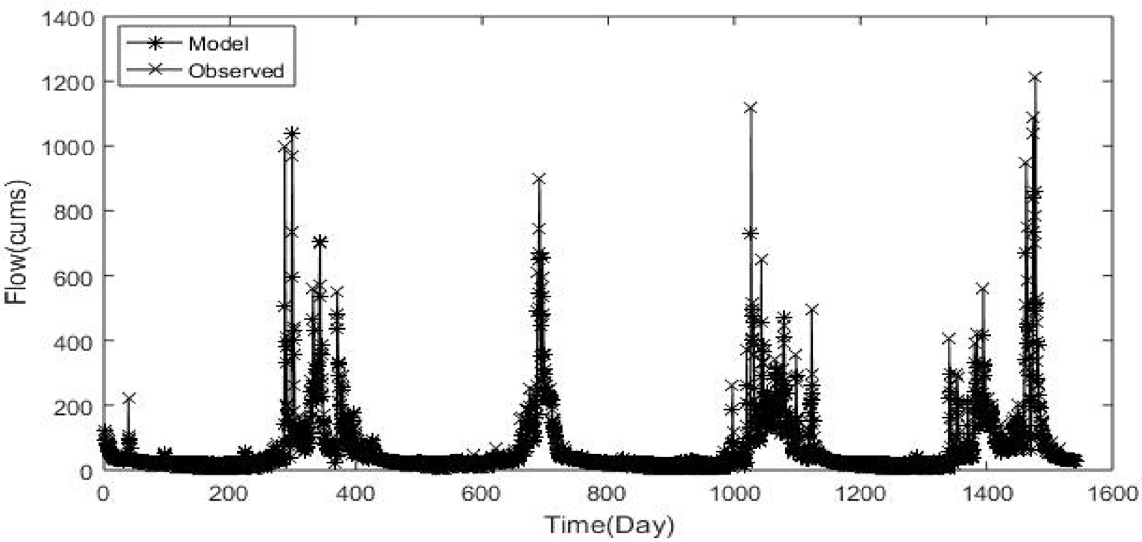


(b)

Fig. 3: Plot of observed and model flow for ANN (a) and WNN (b) model during validation of Kashinagar GD site

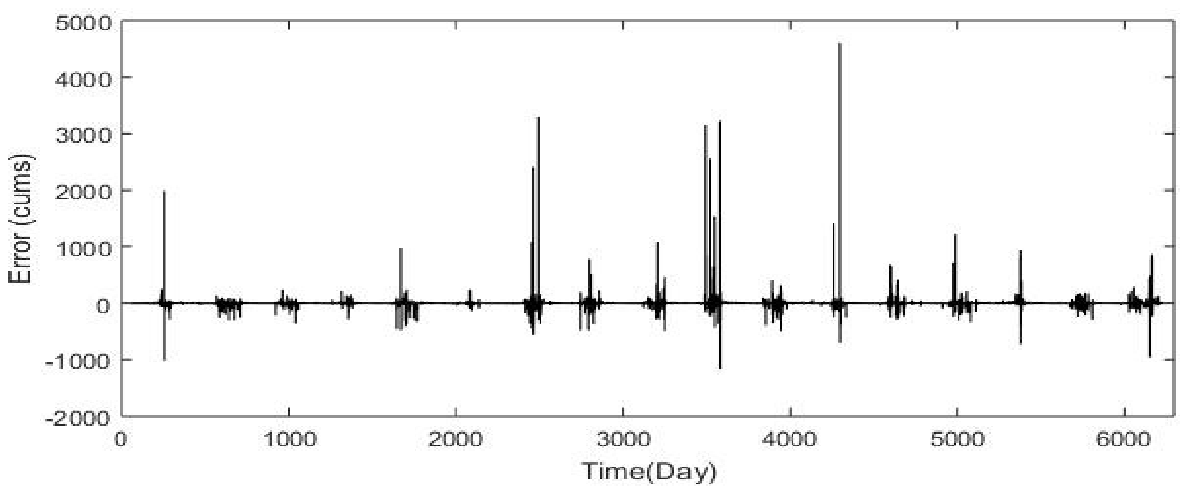


(a)

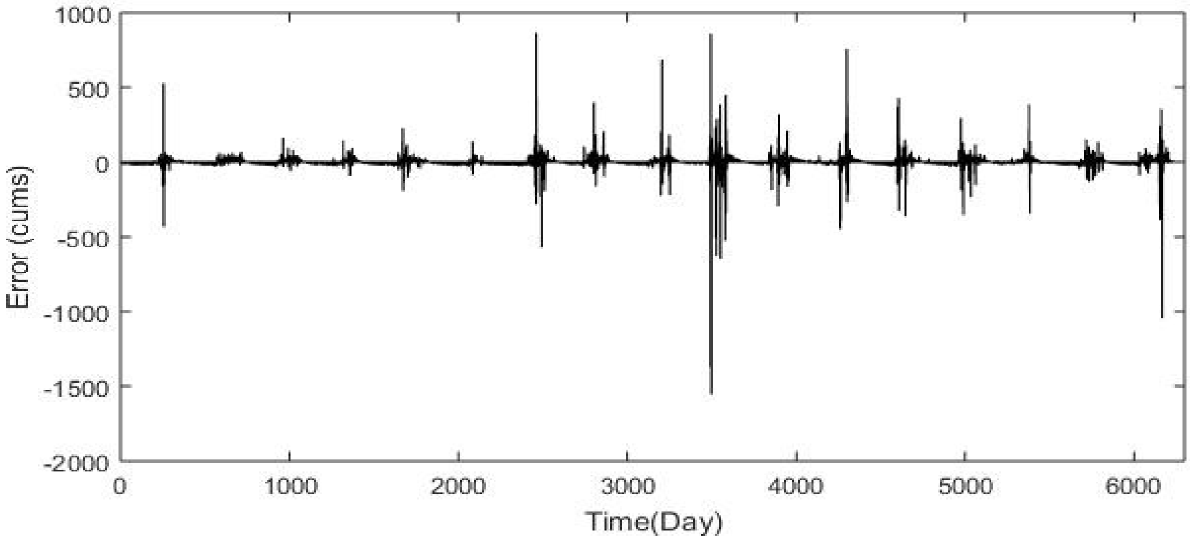


(b)

Fig. 4: Plot of observed and model flow for ANN (a) and WNN (b) model during validation of Gunupur GD site



(a)



(b)

Fig. 5: Distribution of error plots along the magnitude of flow for (a) ANN model and (b) WNN model during validation period

CONCLUSION

This paper reports a hybrid model called wavelet based neural network model for time series modelling of flood forecasting of east flowing river in Andhra region. The proposed model is a combination of wavelet analysis and artificial neural network (WNN). Wavelet decomposes the time series into multi levels of details and it can adopt multi resolution analysis and effectively diagnose the main frequency component of the signal and abstract local information of the time series. The proposed WNN model has been applied to daily rainfall and stream flow of Vamisidhara river basin. The time series data of rainfall and stream flow was decomposed into sub series by DWT. Each of the sub-series plays distinct role in original time series. Appropriate sub-series of the variable used as inputs to the ANN model and original time series of the variable as output. From the current study it is found that the proposed wavelet neural network model is better in forecasting of flood of Vamsadhara river. In the analysis, original signals are represented in different resolution by discrete wavelet transformation; therefore, the WNN forecasts are more accurate than that obtained directly by original signals.

REFERENCES

1. Abrahart R. J. (2012). Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting. Prog Phys Geogr 36(4):480–513.
2. Addison P. S. (2000). The illustrated wavelet transform handbook. Institute of Physics Publishing, London.
3. Alexander. A. A , Santosh G. Thampi, Chithra N. R. (2018). Development of hybrid wavelet-ANN model for hourly flood stage forecasting, Volume 24, Issue 2. doi.org/10.1080/09715010.2017.1422192.
4. Anctil F, Tape D. G,. 2004. An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. Journal of Environmental Engineering Science 3(S1):S121–S128. doi.org/10.1139/s03-071.
5. ASCE Task Committee (2000a). Artificial neural networks in hydrology-I: preliminary concepts. Journal of Hydrologic Engineering 5(2):115–123.
6. ASCE Task Committee (2000b). Artificial neural networks in hydrology-II: hydrologic applications. Journal of Hydrologic Engineering 5(2):124–137.
7. Box, G. and Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.
8. Box, G.E.P, and G.M. Jenkins (1976). Time series analysis: Forecasting and control. Holden-Day. Rev.ed, San Francisco.
9. Cannas B, Fanni A, See L, Sias G (2006). Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Phys Chem Earth, Parts A/B/C 31(18):1164–1171. doi. org/10.1016/j.pce.2006.03.020.
10. Celso A. G. S; Paula K. M. M. F; Richarde M. ; Seyed A. Ai. (2019). Hybrid Wavelet Neural Network Approach for Daily Inflow Forecasting Using Tropical Rainfall Measuring Mission Data. Journal of

Hydrologic Engineering. Volume 24 Issue 2. DOI: 10.1061/(ASCE) HE.1943-5584.0001725.

11. Daubechies I, Bates B. J. (1993). Ten lectures on wavelets. J Acoust Soc Am 93(3):1671–1671.

12. Daubechies, I. (1992). Ten Lectures on Wavelets, vol. 61, SIAM, Philadelphia, Pa, USA.

13. Hecht-Nielsen, R. (1990). Neurocomputing, Reading, MA: Addison-Wesley.

14. Ghafoor, A., & Hanif, S. (2005). Analysis of the Trade Pattern of Pakistan: Past Trends and Future Prospects. Journal of Agriculture and Social Science , 1 (4), 346-349.

15. Mallat, S. (2009). A wavelet tour of signal processing, A Wavelet Tour of Signal Processing,

16. Ramana, R. V., Krishna, B., Kumar, S. R., Pandey, N.G. (2013).Monthly rainfall prediction using wavelet neural network analysis. Journal of Water Resources Manage, 27, 3697-3711.

17. Shoaib M, Shamseldin A. Y, Melville B. W 2014. Comparative study of different wavelet based neural network models for rainfall–runoff modeling. Journal of Hydrology 515:47–58.

18. Shoaib M, Shamseldin A. Y, Melville B. W, Khan M. M. (2015). Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. Journal of Hydrology 527:326–344.

19. Shoaib M, Shamseldin A. Y, Melville B. W, Khan M. M. (2016). Hybrid wavelet neuro-fuzzy approach for rainfall runoff modeling. Journal Comput. Civ. Eng 30(1):04014125. doi.org/10.1061/(ASCE)CP.1943-5487.0000457.

20. Wei S, Yang H, Song J, Abbaspour K, Xu Z. (2013). A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. Journal of hydrological Sciences 58(2):374–389. doi.org/10.1080/02626667.2012.754102.