International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7, Issue-6C2, April 2019

Development of Data Driven Rainfall-Runoff Model for the Sarada River Basin

K.N.V. Rama Devi, R. Venkata Ramana, Y. R. Satyaji Rao and Sanjeet Kumar

Abstract: ***Determining the relationship between rainfall and runoff for a basin is one of the challenging tasks faced by hydrologists and engineers. Conceptual rainfall-runoff models are most suitable in case of data scarcity. However, data driven models are more useful to handle nonlinearity between rainfall-runoff time series data. In this paper an attempt has been made to develop data driven models (Linear and non-linear models) for the Sarada river basin in Vishakhapatnam district of Andhra Pradesh, India. The catchment area of the Sarada river basin is 2665 Sq km. The observed daily rainfall obtained from IMD and daily runoff data obtained from CWC for a period of twenty four years (1989-2013). Autoregressive Integrated Moving Average (ARIMA) linear model and Artificial Neural Network (ANN) and Wavelet Neural Network (WNN) nonlinear models have been developed for the Sarada River basin. The 60% of observed data has been used for calibration and 40% of the data for validation. The comparison of model performance was conducted based upon different statistical indices. The result indicates WNN model performed better than ANN and AIRMA for rainfall-runoff modelling in the Sarada river basin.***

Index Terms: ANN, WNN, Rainfall, Runoff, ARIMA.

I. INTRODUCTION

For an effective and efficient management of the water sources with extensive physiographic variability a contemplative rainfall-runoff relationship modeling adds on as a perquisite[20]. So, as to utilize the privileges attained by the modeling form the past two decades copious linear and non-linear hydrologic models are been developed for the co-relations of rainfall- runoff-sediment in watershed fluvial system. Perpetually these models are trampled to classes as conceptual, lumped, hydraulic and hydrological models, but to do so an ample variety of parameters are required for physically based models as like land use, soil characteristics, soil horizons, treatment of watershed, man-made activities, conservation practices, soil moisture variability, terrain information, roughness of the surfaces etc. Although these parameters vary with the time and space and tedious to monitor on the watershed scale, this led to the insurability of the performance of any good model basing on these parameters in the past. But the recent studies show an improvement in accuracy of the models mainly in time series predictions where the Fuzzy logic, Artificial Neural Networks (ANN) and Genetic Programing (GP) were used. As all these models are data driven models and has tolerance towards uncertainties, imprecision, and fragmentary truth of the input information, this gives a chance to cope up with the drawbacks of the traditional mathematical models and also takes care about hydrological processes. Process-based models apply various physical principles. Whereas when it comes to the ANN like models i.e., black-box driven models map the non linear relationships between input-output parameters without considering the underlying physical process. Therefore, data driven models (ANN) have attained more efficiency and profound consideration and admit from past two decades. The detailed application of ANNs in hydrology and its limitations are discussed by ASCE Task Committee[2, 5]. Utilization of Wavelet Transform (WT) has been discovered successful in managing the non-stationary data[11]. The WT tool improves the performance of data driven models by taking into consideration of both the spectral and temporal information in a given time series data. The WT disintegrates the principle time arrangement information into its sub-parts as details and approximation. Many authors were applied WT to increase forecasting efficiency of ANN[12, 16, 18, 4].

II. STUDY AREA

The catchment area of the Sarada river basin is 2665 Sq km and lies within North latitude 17°25' to 18°17' and East longitude 82°32' to 83°06'. Sarada River is one of the minor river basins that originates near Lakshmipuram village in the Eastern Ghats and joins the Bay of Bengal in the northeastern coastal area of Andhra Pradesh. The typical soils found in the basin are red loamy soils, red sandy soils, coastal sands and alluvial soils. The basin annual rainfall is found to be between 700 to 1000 mm. Most of the rainfall in this region is received during the southwest monsoon than northeast monsoon. Fig.1 shows location of Sarada river basin.

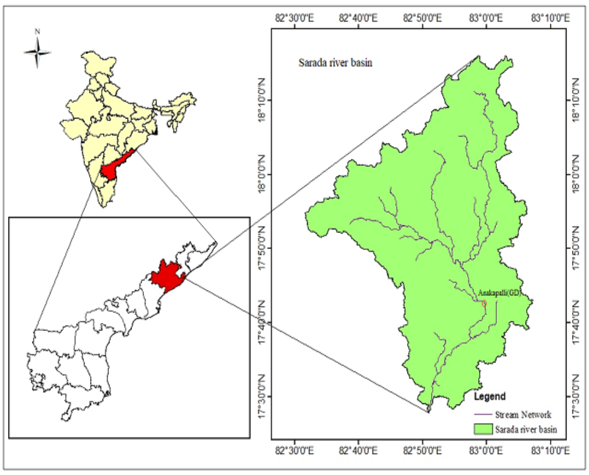

Fig.1: Location of basin study (Sarada River)

Revised Manuscript Received on April 15, 2019.
K. N. V. Rama Devi, Civil Engineering, KLEF, Vijayawada, AP, India.
R.Venkata Ramana, Scientist 'D', NIH, Kakinada, AP, India
Y. R. Satyaji Rao, Scientist 'G', NIH, Kakinada, AP, India
Sanjeet Kumar, Civil Engineering, KLEF, Vijayawada, AP, India

III. METHODOLOGY

ARIMA stochastic models[9] depict 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[9]. The model, "ARIMA" consists of three terms they are i) AR ii) I and iii) MA terms. The "autoregressive (AR)" term indicates Lags of differenced time series in the forecasting equations, whereas "moving average (MA)" term indicates lags of the forecasted errors and the term "Integrated (I)" indicates that the time series which requires distinction to become stationary[1]. AR(p), MA(q) and Auto regressive moving average (ARMA(p,q)) models are some unique instances of Box and Jenkins ARIMA model. In this study ARMA, ARIMA model has been developed for rainfall and runoff modeling of sarada river basin.

A. Artificial neural network (ANN)

In late 1950's a branch of artificial intelligence for posterity was modelled i.e., Artificial Neural Network (ANN) by replicating the neuron networks and their functions in human brain. The developed ANN consists of the parallel-distribution systems which are made of extensive interconnected nonlinear processing elements (PEs) called neurons[13]. Among all others ANNs the Multilayer Perceptron Model (MLP) gained its popularity due to the fast data analysis and information processing ability and due to the availability of suitable hardware these factors led to the inclination towards this ANNs growth from past decade. The MLP is structured of neurons and connections by congruency of neurons arranged in layers, where the input layer i.e., one or more hidden layers and an output layer. The uniqueness of this model network is that the input layer severs as a buffer and distributes the input signals to the hidden layer, and each hidden layer adds up the input, process it with a transfer function, and then distributes the attained results to the output layer which also consist hidden layers possibly connected in the same fashion of input layer and also operates in the same fashion. This makes the model unique and so defied as feed-forward ANNs were the data flows within the network pattern, from one to other layers without any return path.

B. Wavelet Transform

The properties like irregularity in shape and compactness make wavelets an ideal tool for analysis of non-stationary signals. The shifting (delaying) of the mother wavelet provides local information of the signal in time domain, whereas scaling either stretching or compressing of the mother wavelet provides local information of the signal in frequency domain[8]. The shifting and scaling operations applied to mother wavelet are used to calculate wavelet coefficients that provide a correlation between the wavelet and local portion of the signal. From the calculated wavelet coefficients, we can extract two types of components: Approximations (A) and Details (D). Continuous wavelet transform (CWT) is one type of wavelet transform which operates at every scale from that of the original signal up to some maximum scale. Large computational time is required for calculation of wavelet coefficients at every scale. To reduce the computational time, it is preferred to calculate wavelet coefficients for a 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 as DWT[10]. The process of decomposition can be iterated, with successive approximations being decomposed in turn (discarding detail coefficients), so that the original signal is broken down into many lower-resolution components. This process is termed as multi-resolution analysis. We have selected Db3 (length-4 Daubechies) [8] wavelet as mother wavelet, as this is one of the commonly used wavelets for separating fluctuations from the given time series. Based on the number of vanishing moments the smoothness of different wavelets changes [10]. 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 rainfall and runoff time series. The forward discrete wavelet transform is employed to decompose original time series of every variable at different scale (maximum level n = 10). 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 rainfall/runoff variable. The objective behind addition of DW subseries having high correlation with the rainfall-runoff time series is to reduce the number of variables (dimensions or inputs) and to increase the correlation between newly generated subseries and the original rainfall-runoff series. This process is repeated for every set of rainfall-runoff time series data.

C. Selection of mother wavelet function

The selection of the mother wavelet depends on the type of time series data and the performance of each wavelet is very sensitive to the selection of the mother wavelet function. The Daubechies and Morlet wavelet transforms are very popular Mother wavelets even though a number of wavelet families are available. Most popular Daubechies db wavelet for the hydrological applications is db3 or db4[7]. The details of different wavelet families are available in many text books[7, 3].

IV. RESULTS AND DISCUSSION

As per CWC report Sarada river basin is falling in flood prone areas in the sub zone 4 (A) of the Andhra region. Daily gauge discharge data downloaded from India-WRIS for gauge discharge station Anakapalli from 1989–2013 and IMD 0.5 degree gridded daily rainfall of Sarada basin has been used in the analysis. 60% of data is used for calibration and 40% for validation of the model. The performance of various models during testing and training were evaluated by statistical indices namely, the Root Mean Squared Error (RMSE), Correlation Coefficient (R) and Coefficient of Efficiency. The original time series was decomposed into Details (D) and Approximations (A) to a certain number of sub-time series $\{D_1, D_2 \ldots D_p, A_p\}$ by wavelet transform algorithm. Each sub-time series behaves distinctly[17] and contribution of sub-series to original time series varies differently for each successive approximations. So that one signal is decomposed into many lower resolution signal components and they are tested using different scales from 1 to 10 with different sliding window amplitudes. This study deals with a very irregular shaped wavelet signal, the Daubechies wavelet of order 3 (Db3), has been used at level 3. Consequently, D_1, D_2, D_3 where detail time series, and A_3 was the approximation time series[19].

The constructed ANN with the sub-series $\{D_1, D_2, D_3, A_3\}$ at time t are input to ANN and gives an output of original time series at time $t + T_l$, where T_l is length of forecasted time. The WNN input neurons are the sub-series of antecedent values of the precipitation and discharge which were mentioned in Table I. The WNN model was formed in which the weights are learned with Feed forward neural network with Back Propagation algorithm. The numbers of hidden neurons for BPNN are optimized by trial and error method. The performance of various models was determined to predict the river flow and results were mentioned in Table II. To decompose the input rainfall data db3 wavelet function is employed in the present work. The db3 wavelet function with eight vanishing moments has capability to best illustrate the temporal and the spectral information in the input rainfall data[14, 15]. From Table II, it is inferred that low RMSE values of Sarada GD site at Anakapalli (87.0687 m3/s to 48.9945 m3/s) for WNN models when compared to ARMA, ARIMA and ANN models.

Table I: 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

It has been noticed that WNN models predict 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 estimate minimum root mean square error (RMSE) for Sarada basin GD site at Anakapalli (0.8711 m^3/s), high correlation coefficient and highest percentage of coefficient of efficiency (COE) (Sarada>75) during testing. The Model V for Sarada basin GD site of Anakapalli of WNN was selected as the best fit model to develop the rainfall-runoff model for Sarada river basin to forecast flow in one-day advance. Fig.2, Fig.3, Fig.4 and Fig.5 shows the scatter diagram plotted for performance by WNN and ANN during calibration and validation respectively.

Table II: Performance of ARMA, ARIMA, ANN and WNN models in testing and training

Model		Testing			Training		
		RMSE	**R**	**COE**	**RMSE**	**R**	**COE**
ARMA							
		23.8584	0.8477	71.6003	75.3502	0.6813	42.5863
ARIMA							
		26.7862	0.8834	73.6754	79.3242	0.69054	43.1256
ANN							
Model	I	23.3762	0.8528	72.7179	75.8293	0.6619	41.869
Model	II	23.5743	0.851	72.2536	88.7815	0.6128	20.3353
Model	III	21.8142	0.8734	76.242	81.2646	0.5837	33.2717
Model	IV	21.6969	0.8747	76.4968	82.1069	0.5693	31.8989
Model	V	21.0126	0.883	77.9559	71.8458	0.7021	47.8705
Model	VI	21.0095	0.88314	77.9625	65.3786	0.75577	56.8329
Model	VII	21.0419	0.8829	77.8945	80.868	0.5854	33.9902
WNN							
Model	I	14.59	0.9457	89.3723	87.0687	0.7161	23.3597
Model	II	12.2218	0.9623	92.5423	79.0013	0.6127	36.9204
Model	III	9.0993	0.9794	95.8662	73.9154	0.6864	44.7951
Model	IV	7.3412	0.9866	97.3093	79.8172	0.6456	35.6442
Model	**V**	**6.34691**	**0.99**	**97.9888**	**48.9945**	**0.87103**	**75.7576**
Model	VI	5.8926	0.9914	98.2664	84.6235	0.6063	27.6979

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

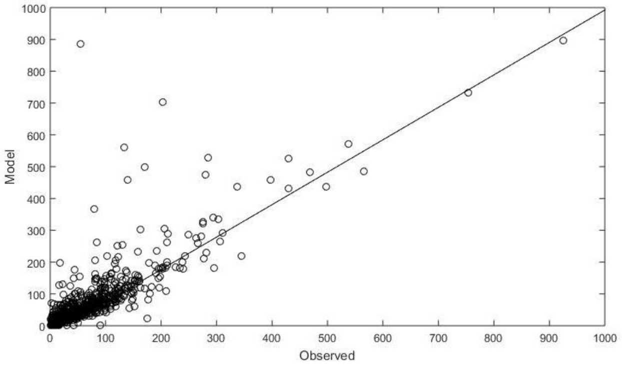

Fig.2: Plot for observed and model Runoff by ANN during training

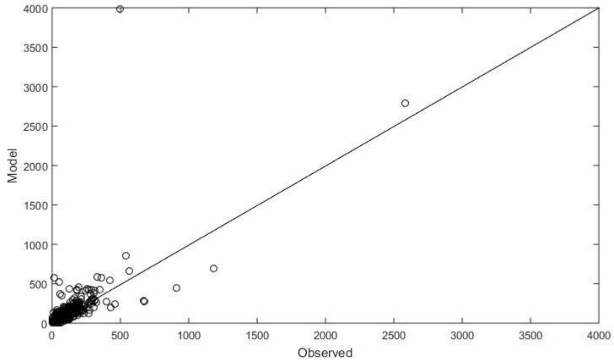

Fig.3: Plot for observed and model Runoff by ANN during testing

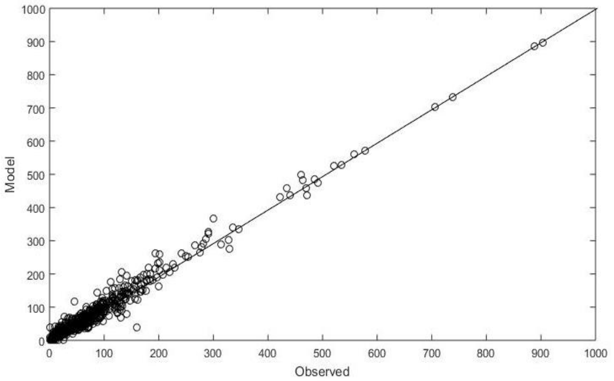

Fig.4: Plot for observed and model Runoff by WNN during training

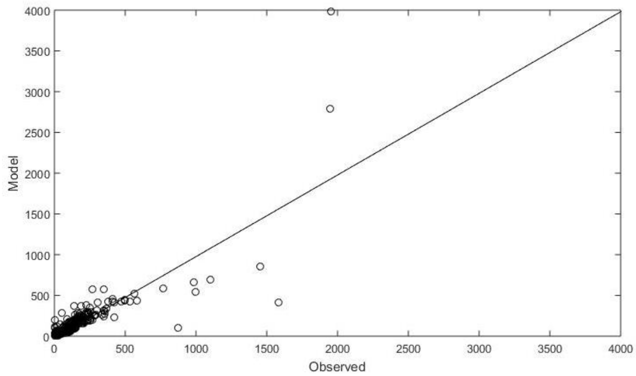

Fig.5: Plot for observed and model Runoff by WNN during testing

Flow predicted by WNN models were very much close to the 45 degree line was noticed from Fig.2, Fig.3, Fig.4 and Fig.5. So based on this investigation, it was worth to point out that the performance of WNN model is high compared to ANN, ARIMA and ARMA models in predicting the river runoff in one day advance.

V. CONCLUSIONS

A hybrid model called wavelet based neural network model has reported this paper for Rainfall-runoff modelling of the Sarada River basin. Combination of wavelet analysis with artificial neural network (WNN) is proposed and it has been applied to daily rainfall and runoff of the Sarada river basin. In this model, discrete wavelet transformation represents original signals in multi resolutions. As the time series data of rainfall-Runoff was decomposed into sub series by DWT. Appropriate sub-series of the variable resolution is used as inputs to the ANN model, so that the performance of model has improved. From the present investigation it was found that the performance of wavelet neural network model is good in forecasting runoff for the Sarada basin in Andhra region. therefore, the WNN model (Model 5) is provided to be a better modelling technique than ARMA, ARIMA and ANN.

ACKNOWLEDGMENT

The authors are grateful to the National Institute of Hydrology, Deltaic Regional Center, Kakinada, India and Koneru Lakshmaiah Education Foundation (Deemed to be University), Andhra Pradesh for collaboration and funding the research.

REFERENCES

1. Abdul ghafoor and sarvet hanif , "Analysis of the Trade Pattern of Pakistan: Past Trends and Future Prospects", *Journal of agriculture & social sciences* (2005), Vol. 01, Issue 04, pp. 346-349.
2. Abrahart R J, Franc¸ois Anctil , Paulin Coulibaly, Christian W. Dawson, Nick J. Mount, Linda M. See, Asaad Y. Shamseldin, Dimitri P. Solomatine, Elena Toth, Robert L. Wilby, "Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting", *Progress in Physical Geography* (2012),Vol. 36, Issue 04, pp. 480–513.
3. Addison PS (2002) the illustrated wavelet transform handbook. Institute of Physics Publishing, London.
4. Anctil F, Tape D G, "An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition", *Journal of Environmental Engineering and Sciences* (2004) , Vol. 03, Issue S1, pp. S121–S128
5. ASCE task Committee, "Artificial neural networks in hydrology ii: hydrologic applications", *Journal of Hydrologic Engineering* (2000),Vol. 05, Issue 02, pp. 124–137

6. Cannas B, Fanni A, See L, Sias G, "Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning" "*Physics and Chemistry of the Earth, Parts A/B/C* (2006)" , Vol. 31, Issue 18, pp. 1164-1171
7. Daubechies I, Bates BJ, " Ten lectures on wavelets", *The Journal of the Acoustical Society of America* (1993), Vol. 93, Issue 03, pp. 1671–1671
8. Daubechies I (1992) Ten Lectures on Wavelets, vol. 61, SIAM, Philadelphia, Pennsylvania, USA.
9. George Box and Gwilym Jenkins (1976), Book on time series analysis.
10. Mallat S.(2009),"A Wavelet Tour of Signal processing"
11. Nason GP, Von Sachs R, "Wavelets in time-series analysis", Philosophical transactions of the Royal Society of London a: mathematical Physical and Engineering Sciences (1999), Vol. 357, Issue 1760, pp. 2511–2526
12. Partal T,"Wavelet regression and wavelet neural network models for forecasting monthly streamflow", *Journal of Water and Climate Change* (2016), Vol. 08, Issue 01, pp.48–61
13. Robert hecht-nielsen (1990) Book on Neurocomputing.
14. Shoaib M, Shamseldin AY, Melville BW, " Comparative study of different wavelet based neural network models for rainfall–runoff modeling", *Journal of Hydrology* (2014), Vol. 515, pp. 47–58
15. Shoaib M, Shamseldin AY, Melville BW, Khan MM "Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach". *Journal of Hydrology* (2015), Vol. 527, pp. 326–344
16. Wei S, Yang H, Song J, Abbaspour K, Xu Z A, "wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows", *Hydrology and Sciences Journal* (2013), Vol. 58, Issue 2, pp. 374–389
17. Wensheng Wang, Jing Ding , "Wavelet Network Model and Its Application to the Prediction of Hydrology", *Nature and Science Journal* (2003), Vol. 1, Issue 1, pp. 67-71
18. Kumar, Sanjeet., Tiwari, M.K., Chatterjee, C., Mishra, A, "Reservoir Inflow Forecasting Using Ensemble Models Based on Neural Networks, Wavelet Analysis and Bootstrap Method", *Water Resources Management Journal* (2015), Vol. 29, Issue 13, pp. 4863- 4883
19. R. Venkata Ramana, B. Krishna, S. R. Kumar, N. G. Pandey. "Monthly Rainfall Prediction Using Wavelet Neural Network Analysis", *Water Resources Management Journal* (2013), Vol. 27, pp.3697–3711.
20. B. Balaji, Sanjeet Kumar. "Dam Break Analysis of Kalyani Dam Using HEC-RAS", International Journal of Civil Engineering and Technology (2018), Vol.9, pp372-380.

AUTHORS PROFILE

K. N. V. Rama Devi, pursuing M.Tech degree in Geoinformatics at K L Deemed to be university, Guntur, Andhra Pradesh, India. Received the B.tech degree in civil engineering from Srinivasa Institute of Engineering & Technology, cheyyeru, Andhra Pradesh.

R.Venkata Ramana, did M.Tech in IIT Madras, presently he is working as a scientist at National Institute of Hydrology, Kakinada, India and his area of interests are urban hydrology, soft computing techniques and applications of RS and GIS.

Dr. Y. R. Satyaji Rao, working as a Scientist at National Institute of Hydrology, Kakinada, India. Graduated in Civil Engineering from Osmania University, Hyderabad. Post graduated in Hydrology and Water Resources Engineering from Anna University, Chennai and PhD from IIT Delhi in the Department of Civil Engineering.

Dr. Sanjeet Kumar, working as Associate Professor, Department of Civil Engineering, Koneru Lakshmaiah Education Foundation Deemed to be University, Andhra Pradesh, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication