

## **Auto Updating Wavelet Based MLR Models for Monsoonal River Discharge Forecasting**

**Vinit Sehgal<sup>1</sup> and Chandranath Chatterjee<sup>2</sup>**

*<sup>1</sup>Department of Civil Engineering, BIT Mesra, Ranchi, INDIA*

*<sup>2</sup>Department of Agricultural and Food Engineering  
Indian Institute of Technology Kharagpur, West Bengal, INDIA.*

### **Abstract**

Development of accurate and reliable flood forecasting models is important for devising the flood fighting and water diversion strategies. The regression based models offer great promise in modeling complex and nonlinear hydrological processes without underlying physics being explicitly provided. In the present work, Auto updating–Wavelet–Multiple Linear Regression (WMLR-AU) hybrid models are developed to forecast monsoonal flood discharge at Naraj Gauging station in Mahanadi basin, India using antecedent time series of water discharge at the same gauging station. The results are presented for 1, 3 and 5 day lead time and are compared with Wavelet–Multiple Linear Regression (WMLR) and Multiple Linear Regression (MLR) models. The WMLR and MLR models are not facilitated with automatic updating and hence are not able to adapt to the changing river discharge patterns effectively. The results show that for 1-day lead time forecasts, both WMLR-AU and WMLR models perform with similar accuracy. For longer lead time, the “Auto Updating” modes outperform WMLR models. MLR models either over predict or under predict in all the cases.

**Keywords:** Wavelet analysis, flood forecasting, Multiple linear regression, Mahanadi basin.

### **1. Introduction**

River discharge forecasting is crucial in water resources management and real-time prediction of extreme floods. Physical models are traditionally utilized to make

discharge forecasts, but these large-scale systems have not proved to be accurate enough to provide timely flood warnings. On comparison, data-based artificial intelligent models have recently gained popularity in hydrological applications due to their rapid development, fewer data requirement and ease of real-time implementation (Adamowski, 2008). Artificial neural networks (ANN) Auto Regression (AR) are examples of most widely used artificial intelligent models. However, these models are found inadequate when dealing with transitory or non-stationary data.

In the last decade, wavelet transform has become a useful technique for analyzing variations, periodicities and trends in a time series Hence it has been used in diverse hydrological problems like drought forecasting (Kim and Valdes, 2003), streamflow analysis (Admowski, 2008; Coulibaly and Burn, 2004; Kucuk and Agiralioglu, 2006; Smith et al 1998), precipitation analysis (Kim, 2004; Lu, 2002; Partal and Kisi, 2007; Xingang et al., 2003), rainfall- runoff relationship (Labat et al., 2000), prediction of river discharge (Zhou et al., 2008); analysis of suspended sediment load (Rajae et al., 2010); estimation of unit hydrographs (Chou and Wang, 2002) and various other hydrological predictions (Wang and Ding, 2003).

In the current study, Auto updating Wavelet- Multiple Linear Regression (WMLR-AU) based models are developed for river discharge forecasting and results are discussed for 1,3 and 5 day lead time. For the purpose of comparison, WMLR and MLR models are also developed for the same lead time.

## 2. Methodology

### 2.1 Wavelet Analysis

The Wavelet analysis is similar to Fourier analysis. In Fourier analysis, signal is broken into sinusoids of unlimited duration, whereas, in wavelet analysis, wavelets are used instead of the sinusoids. Wavelets have waveforms of limited duration with a mean value of zero. In wavelet analysis, the wavelet is shifted forward in steps along full signal. At each step, correlation of wavelet to the signal is measured. When the full series is covered, a set of wavelet coefficients is generated having same consistency in time as that of original signal. The process is repeated. Thus, sets of wavelet coefficients at different scales are generated. The main advantage of using the wavelet method is its robustness since it does not include any potentially erroneous assumptions or parametric testing procedures. Another advantage of the wavelet method is that wavelet variance decomposition allows one to study different investing behavior in different time scales independently. A good illustration on wavelet analysis can be found in the books of Debnath (2010) and Rao (2004).

### 2.2 Multiple Linear Regression

It has been observed that river flows at a particular time step is correlated with flows that precede and succeed it; therefore, the system can generate internal dynamics between inputs and outputs. A  $P^{th}$  order multiple linear regression, i.e. MLR ( $P$ ), refers to the correlation size between values in a time series that are  $P$  periods apart, i.e.

$$x_i = \sum_{i=1}^P \theta_i x_{i-1} + \varepsilon_t \quad (1)$$

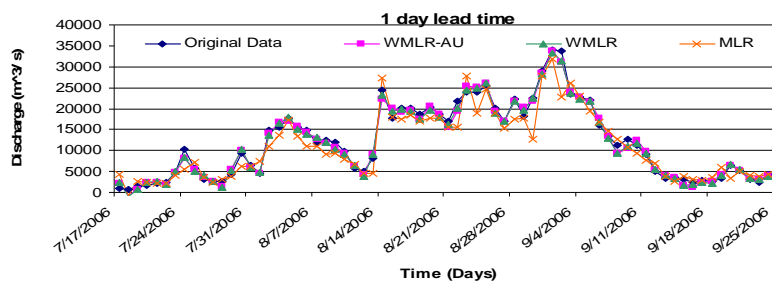
Where,  $\theta_i$  are the MLR coefficients,  $x_i$  is the time series under investigation,  $P$  is the order (length) of the MLR model, and  $\varepsilon_t$ , the residue term, is assumed to be the Gaussian white noise.  $P$  is generally much less than the data length of the series. Thus, in MLR, the current river discharge can be estimated by a linear weighted sum of previous discharges in the series. The weights are the MLR coefficients which are normally estimated using the least square method.

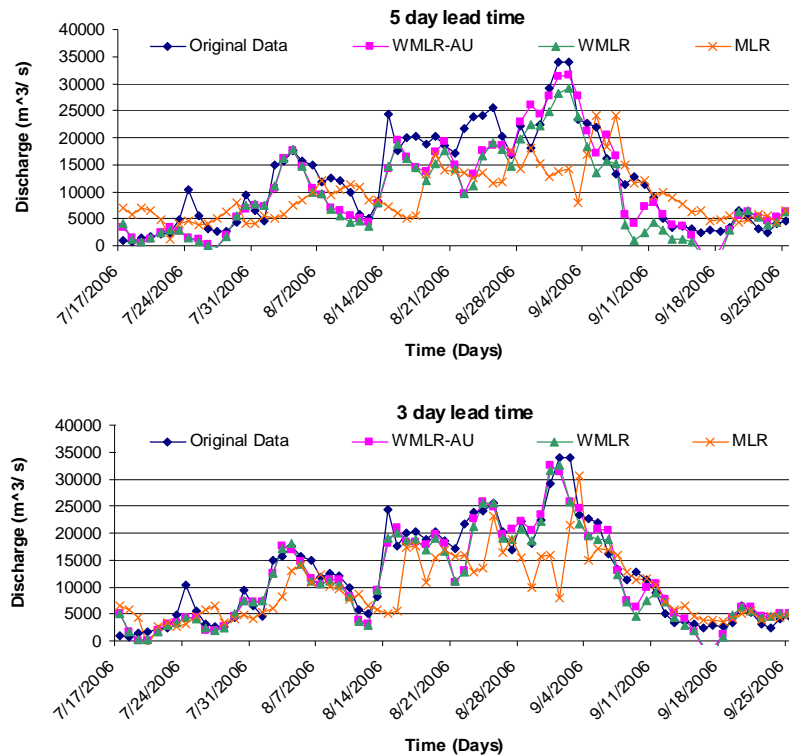
### 3. Focus River

The Mahanadi river basin, one of the largest river basin in India with draining area of 1,41,589 km<sup>2</sup> and main river length of 851 km (Fig.1). Mahanadi River basin lies between 80° 30'E to 86° 50'E longitude and 19° 21'N to 23° 35'N. There are several dams, irrigation projects and barrages present in the Mahanadi river basin with Hirakud reservoir being the most prominent structure in the basin. The main river reach extends from Hirakud Dam to Naraj gauging station, with a total length of 358.4 km. The basin receives most of the rainfall during the monsoon season i.e., between the months of June and September. In the delta region of the Mahanadi river basin flooding is a serious problem during monsoon seasons. Naraj gauge station, situated at the mouth of the delta region, is selected for daily discharge forecasting.

**Table 1:** Performance indices for 1,3 and 5 day lead time forecasts for WMLR-AU, WMLR and MLR models.

1 Day lead	WMLR-AU	WMLR	MLR
RMSE (m <sup>3</sup> /s)	1033.49	941.50	2800.97
MAE (m <sup>3</sup> /s)	794.28	785.07	1959.14
CC	0.994	0.993	0.955
3 Days lead	WMLR-AU	WMLR	MLR
RMSE (m <sup>3</sup> /s)	2778.54	2810.87	6110.92
MAE (m <sup>3</sup> /s)	2062.30	2094.33	4107.38
CC	0.955	0.952	0.767
5 Days lead	WMLR-AU	WMLR	MLR
RMSE (m <sup>3</sup> /s)	4037.21	4667.22	7398.09
MAE (m <sup>3</sup> /s)	2996.06	3523.88	5486.51
CC	0.910	0.903	0.605





**Figure 2:** Observed v/s forecasted discharge for WMLR-AU, WMLR and MLR models.

#### 4. Model Development and Implimentation

Formulated models were derived using 426 monsoonal (June to Oct) water discharge data of the Naraj Gauging station at Mahanadi River for the years 2000-05. After the models were successfully derived, their performance was evaluated using the verification dataset which consisted of another 71 monsoon water discharge data for the year 2006. Table 2 summarizes statistical information on the observed datasets.

**Table 2:** Statistics of the datasets for Daily River flow forecasting (13th July to 25th Sep for the years 2000 to 2006)

Naraj Gauging Site	Training	Testing
Count	426	71
Maximum ( $\text{m}^3/\text{s}$ )	36714.74	33979.18
Minimum ( $\text{m}^3/\text{s}$ )	215.07	696.69
Standard dev. ( $\text{m}^3/\text{s}$ )	7553.05	8737.61
Mean ( $\text{m}^3/\text{s}$ )	6797.52	11979.76
Range ( $\text{m}^3/\text{s}$ )	36499.67	33282.49

A correlation study was carried out to determine the correlation of the current water level time series with antecedent day’s time series. The study revealed that time series upto t-6 days possessed significant information of the current day water level. Inclusion of more antecedent time series would complicate the models with no significant improvement in its predicting ability.

To derive the models, first, the observed river discharge time series was decomposed into sub-time series at three resolution levels by the dyadic DWT. These wavelet sub time series upto lag6, were then used as inputs to Multiple linear regression models (WMLR). The coefficients obtained from the training datasets were applied on the testing datasets to obtain discharge forecasts. In WMLR-AU models, for forecasting discharge on day t+1, data till time t<sup>th</sup> day was taken as training dataset and the coefficients obtained were applied to forecast discharge for t+1<sup>st</sup> day. The process was repeated for every data point in the testing dataset. MLR models are similar in structure as WMLR models and use raw antecedent river discharge data till t-6 days for model formation.

All the developed models were successfully derived and tested with satisfactory performance indices for 1 day advance prediction. The performance of the models was measured by three statistical indices, e.g., the root mean square error (RMSE), the coefficient of correlation (R). They are defined as:

$$RMSE = \frac{\sqrt{\sum (sp - so)^2}}{N} \tag{2}$$

$$CC = \frac{\sum [so - \bar{so}] [sp - \bar{sp}]}{\sqrt{\sum [so - \bar{so}]^2 \sum (sp - \bar{sp})^2}} \tag{3}$$

$$MAE = 1/N(\sum_1^2 |So - Sp|) \tag{4}$$

where, Sp and So are predicted and observed discharge, respectively and N is the number of data points.

It was observed, for 1 day lead time, both WMLR-AU and WMLR models gave similar results. The variance in performance magnified with increase in the lead time. MLR models are observed to show significant deviation from the observed discharge values compared to the other two models. Table 1 shows the performance of WMLR-AU, WMLR and MLR models with 1,3 and 5 days lead time. Figure 2 plots performance of models under study for the testing period. It is evident that Auto Updating algorithm enables the model to adapt with changing discharge pattern of the river and hence deliver reliable and accurate forecasts.

### 5. Conclusion

In this study, Auto Updating Wavelet Multiple Linear Regression models are presented for river discharge forecasting at Naraj Gauging station in Mahanadi basin, India and the results are presented for 1,3 and 5 day lead time. The results indicate superior performance of “Auto Updating” models for higher lead time. For 5 day lead time WMLR-AU gave RMSE of 4037.21m<sup>3</sup>/s compared to 4667.22 m<sup>3</sup>/s and 7398.09 m<sup>3</sup>/s from WMLR and MLR models. WMLR-AU exhibited low MAE and higher CC value compared to the other two models. This validates that Auto updating

methodology enhances the model performance and can prove to be an important tool in real time flood forecasting.

## References

- [1] Adamowski, J. F. (2008), River flow forecasting using wavelet and cross-wavelet transform models. *Hydrological Processes*, 22, 4877-4891.
- [2] Chou CM, Wang R. (2002), On-line estimation of unit hydrographs using the wavelet-based LMS algorithm. *HydroSci J* 47(5) :721–738
- [3] Kucuk M, Agiralioglu N. (2006), Wavelet regression techniques for streamflow predictions. *J Appl Stat* 33(9):943–960.
- [4] Smith LC, Turcotte DL, Isacks B. (1998), Stream flow characterization and feature detection using a discrete wavelet transform. *Hydro Process* 12:233–249.
- [5] Kim S. (2004), Wavelet Analysis of Precipitation Variability in Northern California, USA.” *Jour of Civil Eng KSCE*, , 8, 471- 477.
- [6] Kim, T. W.; Valdes, J. B. (2003), Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J. Hydro. Eng.*, 8 (6), 319-328 .
- [7] Labat, D. (2005), Recent advances in wavelet analyses: Part 1: A review of concepts. *J. Hydro.*, 314 (1-4), 275-288.
- [8] Lokenath Debnath (2010), Wavelet Transforms And Time-Frequency Signal Analysis. *Publisher: Birkhuser Basel, ISBN: 0817641041.*
- [9] Lu RY. (2002), Decomposition of interdecadal and interannual components for North China rainfall in rainy season. *Chin J Atmos (in Chinese)*, 26:611–624.
- [10] Partal, T.; Kisi, O. (2007), Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J. Hydro.*, 342 (1-2), 199-212.
- [11] Rao, G. S. (2004), Wavelet Analysis And Applications. *Publisher: New Age International (p) Limited.*
- [12] Rajae T, Mirbagheri, S. A., Nourani, V. and Alikhani, A. (2010), Prediction of daily suspended sediment load using wavelet and neuro-fuzzy combined model. *Int. J. Environ. Sci. Tech.*, 7(1), 93-110.
- [13] Wang W., & Ding J. (2003), Wavelet network model and its application to the prediction of the hydrology”. *Nature and Science*, 1, 67-71.
- [14] Xingang D, Ping W, Jifan C. (2003), Multiscale characteristics of the rainy season rainfall and interdecadal decaying of summer monsoon in North China. *Chin Sci Bull* 48:2730–2734.
- [15] Zhou HC, Peng Y, Liang G-H. (2008), The research of monthly discharge predictor-corrector model based on wavelet decomposition. *Water Resour Manage* 22(2):217–227.