

## Development of Autoregressive Time Series Model for Prediction of Rainfall and Runoff in Toriya Watershed

ARPAN SHERRING, MANISH DWIVEDI AND A. K. MISHRA

*Department of Soil, Water, Land Engineering and Management, Allahabad Agricultural Institute-Deemed University  
 Allahabad 211007, India  
 E-mail : asherring@rediffmail.com*

### Abstract

A study was conducted to develop a stochastic autoregressive time series model for prediction of annual rainfall and runoff in Toriya watershed located at Lalitpur district of Uttar Pradesh. The study area lies between 24°45' N latitude and 78°20' E longitude and has an area of 106 ha. The developed model is based on 14 years (1994—2007) annual rainfall and runoff data. Autoregressive (AR) model of order 0, 1 and 2 proposed earlier were tried. The goodness of fit and adequacy of model were tested by Box-Pierce Portmonteau test, Akai information criterion (AIC) and various statistical characteristics viz. mean forecast error (MFE), mean absolute error (MAE), mean relative error (MRE), root mean square error (RMSE) and integral square error (ISE). Based on the results, it was concluded that AR (1) model can be effectively used for prediction of rainfall and runoff in Toriya watershed of Lalitpur district.

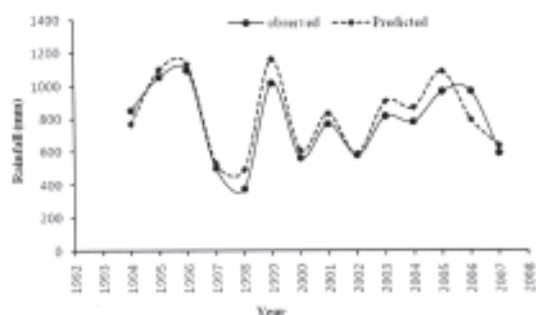
**Key words :** Autoregressive, Time series model, Rainfall, Runoff, Toriya watershed.

Water is essential for sustaining all forms of life, food production, economic development, and for general well being. It is impossible to substitute for most of its uses, difficult to depollute, expensive to transport, and it is truly a unique gift to mankind from nature. Water is also one of the most manageable natural resources as it is capable of diversion, transport, storage and recycling. Water has become a global crisis and it forms the basis for life on earth. Growing population, rapid urbanization and increasing industrial and agricultural growth have increased the pressure on water resources. At the same time climatic changes are affecting the climatic events like rainfall globally and will probably lead to even higher temperatures and lower rainfall in tropical areas. Land degradation due to erosions is caused by nu-

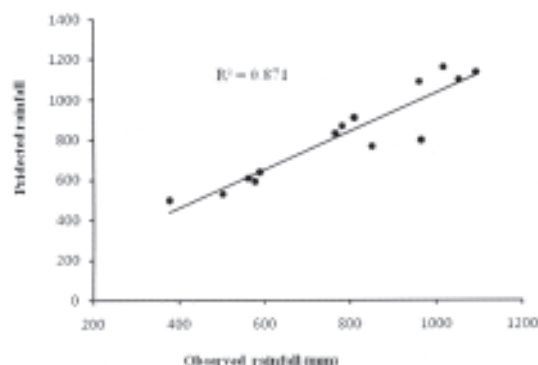
merous factors. Among these, runoff water is the main key factor responsible for erosion. Rainfall runoff modeling is an important area of hydrologic studies and is one in which research is still being actively carried out. Spatially conceptual models only represents the physical processes while the mathematical models only consider the mathematical relationship between the rainfall and runoff without considering the physical processes. These are two broad categories of rainfall runoff models. The reason is that many hydrologic phenomena are highly erratic, complex and random in nature and hence they can be interpreted only in the probabilistic sense. One of the important problems in hydrology deals with interpreting a past record of hydrologic events in terms of future probabilities of occurrence. Hydrologic modeling using

**Table 1.** Statistical parameters of autoregressive (AR) model for annual rainfall in Toriya watershed of Lalitpur district.

Model	Autoregressive parameters	White Noise variance $\sigma_e^2$	Akaike information criterion, AIC (p)	Value of Portemonteau test (Q)	Degree of freedom upto 5 lags	Table value of $\chi^2$ at 5% level of significance
AR (0)	—	51238.84	151.82	0.031	5	11.07
AR (1)	$\Phi_1 = -0.3410$	56093.18	155.09	0.0196	4	9.48
AR (2)	$\Phi_1 = -0.2636$ $\Phi_2 = -0.1149$	50193.92	155.53	1.02	3	7.81



**Figure 1.** Comparison of correlograms of observed and predicted series for annual rainfall of Toriya watershed of Lalitpur district.



**Figure 2.** Comparison between observed and predicted annual rainfall of Toriya watershed of Lalitpur district.

climatological data as a stochastic process is of interest to a variety of hydrological areas such as flood routing, reservoir operation and agricultural planning. The most reliable and authentic type of climatic data required for hydrological model is observed records of rainfall and runoff. It is essential to have a sufficiently long observed data to achieve an operating mathematical rule. Records, which consist of short term data are not suitable for proper planning and arrangement processes and also these consist of only one time realization. Long term climatic data are often not available for a location of interest. In such circumstances, it becomes necessary to extrapolate both deterministic and stochastic component, but stochastic time series models, such as autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) are widely used to predict the annual runoff. The lack of complete understanding of physical processes involved, the consequent uncertainties in the occurrence of future events highlighted the stochastic modeling. These models are suitable for prediction of annual, monthly or daily

rainfall and short term prediction. Iyengqr (1) and Rai and Sherring (2) also reported that a developed autoregressive model is suitable for a certain range and applicable to a particular zone of climate. Keeping this in view an attempt was made to develop a stochastic time series model for prediction of rainfall and runoff for Toriya watershed of Lalitpur district.

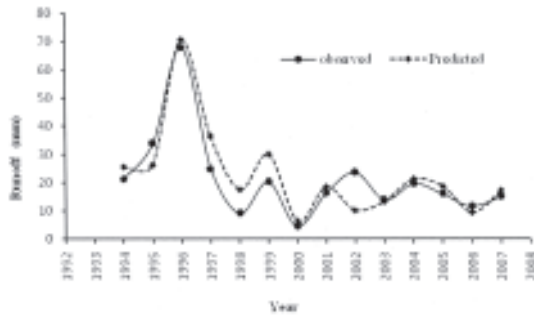
## Methods

### *Description of the Study Area*

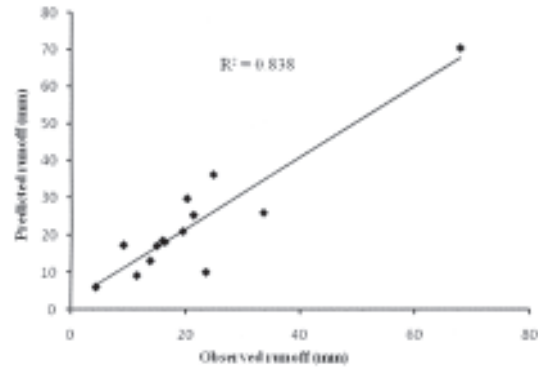
The Toriya watershed, located in the Lalitpur district of the state of Uttar Pradesh, India falls between  $24^{\circ}45' N$  latitude and  $78^{\circ}20' E$  longitude. The catchment area has an elongated shape and is spread over an area of 106 ha. The slope of the catchment is nearly level to gentle and ranging between 1–3%. The average maximum and minimum temperatures are  $44.8^{\circ}C$  in June and  $5^{\circ}C$  in January. The average annual rainfall is 760.5 mm and about 75% of the

**Table 2.** Statistical parameters of autoregressive (AR) model for runoff in Toriya watershed of Lalitpur district.

Model	Autoregressive parameters	White Noise variance	Akaike information criterion, AIC (p)	Value of Porte-monteaue test (Q)	Degree of freedom upto 5 lags	Table value of $\chi^2$ at 5% level of significance
AR (0)	–	249.32	77.26	0.6137	5	11.07
AR (1)	$\Phi_1 = -0.3496$	225.30	77.71	0.4628	4	9.48
AR (2)	$\Phi_1 = -0.3341$ $\Phi_2 = -0.0426$	193.48	77.85	0.5325	3	7.81



**Figure 3.** Comparison of correlograms of observed and predicted runoff for Toriya watershed of Lalitpur district.



**Figure 4.** Comparison between observed and predicted runoff of Toriya watershed of Lalitpur district.

annual rainfall is received during June to September and rest 25% in the remaining period. The daily rainfall (mm) and runoff (ha-m) data of last 15 years (1994—2007) were collected from Soil Conservation Department, Lalitpur, Uttar Pradesh, India.

*Stochastic Time Series Model*

A mathematical model representing stochastic process is called stochastic time series model. A sample time series model could be represented by simple probability distribution  $f(X : \emptyset)$  with parameters  $\emptyset = (\emptyset_1, \emptyset_2, \dots)$  valid for all positions  $t = 1, 2, \dots, N$  and without any dependence between  $X_1, X_2, \dots, X_n$ .

A time series model with dependence structure can be formed as

$$\epsilon_t = \phi \epsilon_{t-1} + \eta_t \quad \dots (1)$$

where  $\eta_t$  is an independent series with mean zero and variance  $(1 - \phi^2)$ ,  $\epsilon_t$  is dependent series and  $\phi$  is the parameter of the model.

Time series modeling can be organized in five stages i.e. identification of model composition, selection of model type, identification of model form, estimation of model parameters and testing of goodness of fit for validation of the model (3).

*Autoregressive (AR) Model*

In the Autoregressive model, the current value of a variable is equated to the weighted sum of a pre-

assigned no. of part values and a variate that is completely random of previous value of process and shock. The  $p^{\text{th}}$  order Autoregressive model, AR (p) is estimated by following equation.

$$Y_t = \bar{Y} + \sum_{j=1}^p \Phi_j (Y_{t-j} - \bar{Y}) + \epsilon_t \quad \dots (2)$$

Where,  $Y_t$  is the time dependent series (variable),  $\epsilon_t$  is the time dependent series which is independent of  $Y_t$  and is normally distributed with mean zero and variance  $\sigma_{\epsilon}^2$ ,  $\bar{Y}$  is the mean annual flow and rainfall data and  $\Phi_1, \Phi_2, \dots, \Phi_p$  are the Autoregressive parameters.

*Estimation of Autoregressive Parameter ( $\Phi$ )*

$$AR(1) : \Phi_1 = \frac{D_{1,2}}{D_{2,2}} \quad \dots (3)$$

$$AR(2) : \Phi_1 = \frac{D_{1,2} D_{3,3} - D_{1,3} D_{2,3}}{D_{2,2} D_{3,3} - D_{2,3}^2} \quad \dots (4)$$

$$\Phi_2 = \frac{D_{1,3} D_{2,2} - D_{1,2} D_{2,3}}{D_{2,2} D_{3,3} - D_{2,3}^2} \quad \dots (5)$$

For estimation of model parameters, method of maximum likelihood was used (4).

**Table 3.** Statistical characteristics of observed and predicted rainfall and runoff for Toriya watershed of Lalitpur district.

	Statistical characteristics	Rainfall (mm)		Runoff (mm)	
		Observed	Predicted	Observed	Predicted
1	Mean	728.70	785.87	17.58	19.08
2	Standard deviation	221.01	233.13	15.29	16.01
3	Skewnss	-0.0396	0.2021	2.41	2.19
4	Mean forecast error	-	-56.35	-	-1.64
5	Mean absolute error	-	56.35	-	1.64
6	Mean relative error	-	-1.087	-	-1.308
7	Root mean square error	-	210.85	-	6.14
8	Integral square error	-	0.074	-	0.677

$$Di, j = \frac{N-j}{N+j-i-j} \sum_{i=0, j=0}^{N+i-(i+j)} Z_i + iZ_j + l \quad \dots (6)$$

Where D is the difference operator, n is the sample size, i and j maximum possible order and l is the autocorrelation function.

The autocorrelation function  $r_k$  of lag k was estimated as proposed by Kottegoda and Horder (5).

$$r_k = \frac{\sum_{t=l}^{N-K} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=l}^N (Y_t - \bar{Y})^2} \quad \dots (7)$$

Where  $r_k$  is the autocorrelation function of time series ( $Y_t$ ) at lag k,  $Y_t$  is stream flow series (observed data),  $\bar{Y}$  is mean of time series  $Y_t$ , N are the total number of discrete values of time series ( $Y_t$ ). The autocorrelogram was used for identifying the order of the model for given time series as well as for comparing the sample correlogram with model correlogram. The 95% probability levels for the autocorrelation function was estimated (6).

*Partial Autocorrelation Function*

The partial autocorrelation  $PC_{kk}$  was calculated

to identify both the type and order of the model (7).

$$PC_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} PC_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} PC_{k-1,j} r_j} \quad \dots (8)$$

Where  $PC_{kk}$  is the partial autocorrelation function at lag K and  $r_k$  is the autocorrelation function at lag K. The 95% probability limit for partial autocorrelation function was also calculated (6).

*Parameter Estimation of AR (p) Models*

The mean ( $\bar{Y}$ ) and variance ( $\sigma_\epsilon^2$ ) of the observed data  $Y_t$  were estimated. After computation of  $\bar{Y}$  and  $\sigma_\epsilon^2$ , the remaining parameters  $\Phi_1, \Phi_2, \dots, \Phi_p$  of the AR models were estimated by

$$Z_t = Y_t - \bar{Y}, \text{ for } t = 1, 2, \dots, N \quad \dots (9)$$

The parameters  $\Phi_1, \Phi_2, \dots, \Phi_p$  were estimated by solving the p system of linear equations (8).

*Statistical Characteristics and Goodness of Fit of Autoregressive (AR) Model*

The MFE, MAE, MRE, RMSE and ISE were estimated to evaluate the performance of the model. The goodness of fit tests of AR model fitted to annual hydrologic series were accomplished by testing the residuals of a dependence model for correlation and the order of the fitted model is compared with the order of the dependence model. The Box-Piece Portmonteau lack of fit test and Akaike Information Criterion (9) was done for checking the order of the fitted model and its adequacy.

**Results and Discussion**

*Model Identification*

The standardize annual rainfall and runoff series were modeled through the autoregressive mod-

els. The modeling procedure of the data series involved various steps like preliminary analysis and identification, estimation of model parameters and diagnostic checking for the adequacy of the model (10). Autocorrelation and partial autocorrelation are used for identification of the proper type and order of the model. The autocorrelation functions and partial autocorrelation functions were determined for the 95% probability limits. The autocorrelation function and partial autocorrelation functions with 95% probability limits upto 5 lag of the series (lag k) were computed and the autoregressive model of first order AR (1) was selected for further analysis.

#### *Models of Autoregressive (AR) Family*

The parameters of AR model were computed for annual rainfall and runoff (Tables 1—3 ; Figs 1—4). The predicted values of annual rainfall and runoff were compared with the observed values. It was observed that AR(p) model upto order 2 has shown the good fit and correlation between the observed and predicted values and given in Figs. 1 and 3.

AR(p) models for prediction of annual rainfall :

$$\text{AR (1) : } Y_t = 760.75 + 0.3410 (Y_{t-1})$$

$$\text{AR (2) : } Y_t = 760.75 + 0.2636 (Y_{t-1})$$

AR(p) models for prediction of annual runoff :

$$\text{AR (1) : } Y_t = 21.20 + 0.3496 (Y_{t-1})$$

$$\text{AR (2) : } Y_t = 21.20 + 0.3341 (Y_{t-1})$$

#### *Box-Pierce Portmonteau and Akaike Information Criterion Test for AR Model*

The Box-Pierce Portmonteau lack of fit test and Akaike Information Criterion (AIC) for all three models to check the adequacy for both annual rainfall and runoff for AR (0), AR (1) and AR (2) models were estimated. The results of Box-Pierce Portmonteau lack of fit test revealed that all three models viz. AR (0), AR (1) and AR (2) were giving good fit and were ac-

ceptable but AIC values of AR (1) in both rainfall and runoff are lying between AR (2) and AR (0), therefore AR (1) was considered for further prediction of annual rainfall and runoff.

#### *Comparison of the observed and Predicted Annual rainfall and Runoff*

The correlogram of observed and predicted series for annual rainfall and runoff were developed by plotting autocorrelation functions against lag K. A graphical comparison of observed and predicted annual rainfall and runoff with the selected model are shown in Figures 2, 4. The graphical representation of the data shows a closer agreement between observed and predicted annual rainfall and runoff by selected model. It reveals that developed model for rainfall and runoff can be utilized for the prediction of future trends with minimum chance of error.

#### *Statistical Characteristics of Data*

The mean, standard deviation and skewness of observed and predicted data were calculated to evaluate the fitting of the model in moment preservation. The results are tabulated in Table 3. The results show that the skewness of predicted data by AR (1) model and observed data are lying between -1 to + 1 and therefore AR (1) model preserved the mean and skewness better.

#### *Performance Evaluation of AR (1) Model for Prediction of Annual Rainfall and Runoff*

Performance of the model was also estimated by estimating statistical characteristics such as MFE, MAE, MRE, MSE, RMSE, ISE to prove the adequacy of the model for future prediction with higher degree of correlation to previously observed observations. In rainfall and runoff, the MFE for AR (1) model is 56.35 mm and 1.64 mm respectively. The different errors between the observed and predicted values of annual rainfall and runoff by AR (1) model were estimated and which represents that for prediction of annual rainfall and runoff by AR (1) model is giving

the best results. Since all the errors are comparatively less and graphical comparison of observed and predicted values of rainfall and runoff indicates that the AR (1) model can be used efficiently for prediction of annual rainfall and runoff in Toriya watershed.

### *Conclusion*

On the basis of the estimated errors, statistical characteristics and correlation between observed and predicted values, it is concluded that the proposed autoregressive AR (1) model can be used to predict the annual rainfall and runoff in Toriya watershed of Lalitpur district.

### **References**

1. Iyengar R. N. 1982. Stochastic modeling of monthly rainfall. *J. Hydrolog.* 57 : 375—387.
2. Rai E. and A. Sherring. 2007. Development of autoregressive time series model for prediction of rainfall and runoff for Mansahara watershed of lower Gomti catchment. *J. Agric. Engg.* 44 : 38—42.
3. Salas J. D. and R. A. Smith. 1981. Physical basis of stochastic models of annual flows. *Water Resour. Res.* 17 : 428—430.
4. Box G. E. P. and G. Jenkins. 1970. *Time series analysis, forecasting and control*. First Ed. Holden-Day Inc., San Francisco.
5. Kottegoda N. T. and M. A. Horder. 1980. Daily flow model rainfall occurrence using pulse and transfer function. *J. Hydrol.* 47 : 215—234.
6. Anderson R. L. 1942. Distribution of the serial correlation coefficients. *Ann. Math. Stat.* 13 : 1—13.
7. Durbin J. 1960. The fitting of time series model. *Rev. Int. Inst. Stat.* 28 : 23—33.
8. Yule G. V. 1927. On a method of investing periodicities in disturbed series with special reference to Wolfer's sunspot number. *Phil. Tran. A* 226 : 267.
9. Akaike H. 1974. A new look at the statistical model identification on *IEEE transactions on automatic control*, AS-19, pp. 716—723.
10. Sharma A., R. V. Singh, S. M. Narulkar, and P. K. Dashora. 2003. Identification of appropriate stochastic model for forecasting monsoonal river inflows. *J. Agric. Engg.* 12 : 191—208.