Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

Shaonlee Chakraborty 1, D.M. Denis 2 and A. Sherring 3
1,2 Department of Soil & Water Conservation Engineering
3 M.Tech (Soil and Water conservation engineering)
2 Professor
3 Associate Professor
1 Email:- shaonleec@gmail.com

Abstract: A study was conducted to develop a stochastic time series model, capable of prediction of rainfall and runoff in Kelo watershed. The Kelo watershed of Mahanadi river basin situated in Raigarh region of Chhattisgarh state. Raigarh is in the North–Central part of Chhattisgarh state. The catchment area of the watershed is 1175 sq km. It is situated 21°53’00’’ latitude and longitude 83°24´00´´. The rainfall and runoff data of the watershed from the year (2001-2006) was collected and used for the development of model. Autoregressive (AR) models of order 0,1 and 2 were tried. The goodness of fit and adequacy of models were tested by Box-Pierce Portmanteau test, Akaike information Criterion and by comparison of historical and generated data correlogram. The graphical representation between historical and generated correlogram, where in both the cases of rainfall and runoff there was a very close agreement between them. The comparison between the measured and predicted rainfall and runoff by AR(1) model clearly shows that the developed model can be used efficiently for the future prediction of rainfall and runoff in Kelo watershed.

Keywords - Stochastic time series model, Autoregressive (AR) models, Akaike information Criterion, Box- Pierce Portmanteau test

I. INTRODUCTION

India is an agrarian country, due to its favourable monsoon climate and vast area of fertile culturable land. The Indian has been traditionally dependent on agriculture as 70% of its population is engaged in farming. The efficient management of water and other natural resources will likely be an significant issue. Rainfall and runoff modeling is an important area of hydrological studies and is one in which research is actively carried out. Methodology for coupling stochastic models of hydrological process applying two different time scales so that the time series generated by different models be consistent. (Koutosoyiannis 2001) Specially conceptual models, which attempt to represent the physical process which occurs on the catchments and mathematical models, which only considered the mathematical relationship between rainfall and runoff without considering the physical process. The principal aim of time series analysis to describe the history of moments in time of some variable at a particular site. A comprehensive review on time series analysis technique used in climatology and hydrology. It was suggested to use more important powerful test for stationary and trend detection in time series. (Machiwal and Jha 2006). Most hydrologic system have both deterministic as well as stochastic component, but stochastic time series model such as Autoregressive (AR) (Thomas and Fiering, 1962; Yevjevich, 1963; Matalas, 1967), moving average (MA) and Autoregressive moving average (ARMA) (Carlson, 1970) are widely used to predict annual runoff. Identification generally depends on the characteristics of overall water resources system, the characteristics of time series and the models input. Salas and Smith (1981) demonstrated these of physical consideration of the type of model. A developed Auto regressive model is suitable for a certain range and applicable for particular zone of climate. (Rai and Sherring 2007). Multivariate autoregressive (MAR) and univariate autoregressive (AR) methods applied to regional scale rainfall runoff modeling (Tomasz 2006). Almost every model available fits in a data analyzed only realistically in a certain range. It is possible to maintain reliable flow simulation by cascading a series of runoff prediction regression model that predict a downstream flow from an upstream flow and the incremental rainfall between gauging station. (Vongtanaboon et.al, 2008). To explore the influence of the inflow on the outflow in a river system and to exploit the internal interaction of the outflows, bivariate time series models were needed. (Wonga et al. 2007). In addition to this it may be noted that models
Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

found applicable in a particular zone eg. the temperate zone not to be applicable in other zone such as tropic. (Iyengar, 1982).

The present study develop a stochastic model of annual rainfall and runoff time series applicable for Kelo watershed in Chhattisgarh state has been planned with following objectives:

1. To generate or develop stochastic time series model for prediction of rainfall and runoff for Kelo watershed.
2. To estimate parameters of Autoregressive model.
3. To test the validity of the predicted rainfall and runoff with measured and evaluated the performance of the developed model.

II. MATERIALS AND METHODS

The study area is Kelo watershed of Mahanadi river basin situated in Raigarh region of Chhattisgarh state as depicted in figure 1.

Fig 1 Location of study area

A. Stochastic Time Series Model

A mathematical model representing stochastic process is called stochastic time series model. It has a certain mathematical form or structure and a set of parameters. A simple time series model could be represented by a single probability distribution function $f(X; \Theta)$ with parameters $\Theta = (\Theta_1, \Theta_2, \ldots)$ valid for all positions $t = 1, 2, \ldots, N$ and without any dependence between $X_1, X_2, \ldots, X_N$.

A time series model with dependence structure can be formed as:

$$\varepsilon_t = \phi \varepsilon_{t-1} + \eta_t \quad \text{---------- (1)}$$

where,

- $\eta_t$ = An independent series with mean zero and variance $(1 - \phi^2)$
- $\varepsilon_t$ = Dependent series
- $\phi$ = Parameter of the model.

(a) 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<sup>th</sup> order autoregressive model AR (p), representing the variable \( Y_t \) is generally written as

\[
Y_t = \overline{Y} + \sum_{j=1}^{p} \Phi_j (Y_{t-j} - \overline{Y}) + \varepsilon_t \quad \text{---- (2)}
\]

Where,
- \( Y_t \) = The time dependent series (variable)
- \( \varepsilon_t \) = The time independent series which is independent of \( Y_t \) and is normally distributed with mean zero and variance \( \sigma^2 \)
- \( \overline{Y} \) = Mean of annual rainfall and runoff data
- \( \Phi_1, \Phi_2, \ldots, \Phi_p \) = Autoregressive parameter.

**B. Estimation of Autoregressive parameter (\( \Phi \)) maximum likelihood estimate**

For estimation of the model parameter method of maximum likelihood will be used (Box and Jenkins, 1970).

Consider the sum of cross-products,

\[
z_j z_k + z_{j+1} z_{k+1} + \ldots + z_{N+1-j} z_{N+1-k}
\]

and define

\[
D_{ij} = D_{ji} = \frac{N}{N+2-i-j} \quad \text{-----(3)}
\]

where,
- \( D \) = difference operator
- \( N \) = sample size
- \( i, j \) = maximum possible order

in particular,

\[
AR (1) : \Phi_1 = \frac{D_{1,2}}{D_{2,2}} \quad \text{-----(4)}
\]

\[
AR (2) : \Phi_1 = \frac{D_{1,2} D_{3,3} - D_{1,3} D_{2,3}}{D_{2,2} D_{3,3} - D_{2,3}^2} \quad \text{-----(5)}
\]

\[
\Phi_2 = \frac{D_{1,2} D_{2,2} - D_{1,2} D_{2,3}}{D_{2,2} D_{2,3} - D_{2,3}^2} \quad \text{-----(6)}
\]

(a) Autocorrelation function

The autocorrelation function \( r_k \) of the variable \( Y_t \) of equation (3.2) is obtained by multiplying both sides of the equation (3.2) by \( Y_{t+k} \) and taking expectation term by term. The relationship proposed by Kottegoda and Horder (1980) for the computation of autocorrelation function of lag \( K \) was used which is expressed as:
Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

\[ r_k = \frac{\sum_{t=1}^{N-K} (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})} {\sum_{t=1}^{N} (Y_t - \overline{Y})^2} \]  

--- (7)

Where,

\[ r_k = \text{Autocorrelation function of time series } Y_t \text{ at lag } k \]
\[ Y_t = \text{Stream flow series (historical data)} \]
\[ \overline{Y} = \text{Mean of time series } Y_t \]
\[ k = \text{Lag of } K \text{ time unit} \]
\[ N = \text{Total number of discrete values of time series } Y_t \]

The following equation was used to determine the 95 per cent probability levels (Anderson, 1942).

\[ r_k (95\%) = \frac{-1 \pm 1.645 \sqrt{N - K - 1}} {N - K} \]  

--- (8)

where, \( N = \text{Sample size} \).

(b) Partial Autocorrelation function

The following equation was used to calculate the partial autocorrelation function of lag \( K \). (Durbin, 1960).

\[ PC_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} PC_{k-1,j} r_{k-j}} {1 - \sum_{j=1}^{k-1} PC_{k-1,j} r_j} \]  

--- (9)

where,

\[ PC_{k,k} = \text{Partial autocorrelation function at lag } K \]
\[ r_k = \text{Autocorrelation function at lag } K \]

The 95 percent probability limit for partial autocorrelation function was calculated using the following equation

\[ PC_{k,k} (95\%) = \frac{1.96} {\sqrt{N}} \]  

--- (10)

(c) Parameter estimation of AR (p) models

The average of sequence \( Y_t \) was computed by following equation:

\[ \overline{Y} = \frac{1} {N} \sum_{t=1}^{N} Y_t \]  

--- (11)

The average \( \sigma^2 \) of \( Y_t \) was computed by the following equation:

\[ \sigma^2 = \frac{1} {(N-1)} \sum_{t=1}^{N} (Y_t - \overline{Y})^2 \]  

--- (12)

After computation of \( \overline{Y} \) and \( \sigma^2 \), the remaining parameters \( \Phi_1, \Phi_2, \ldots, \Phi_p \) of the AR models were estimated by using the sequence:

\[ Z_t = Y_t - \overline{Y} \]  

--- (13)
The parameters $\Phi_1, \Phi_2, \ldots, \Phi_p$ were estimated by solving the $p$ system of following linear equations (Yule and Walker equation):

$$r_k = \Phi_1 r_{k-1} + \Phi_2 r_{k-2} + \ldots + \Phi_p r_{k-p} \quad K>0$$

or

$$\text{orr}_k = \sum_{j=1}^{p} \Phi_j r_{k-j} \quad (14)$$

Where, $r_1, r_2$ were computed from equation 7

### C. Statistical characteristics

#### (a) Mean Forecast Error

Mean forecast error was calculated to evaluate the performance of auto regressive models fitted to time series of rainfall and runoff. The mean forecast error (MFE) was computed for the annual stream flow series by the following equation: (Raghuwanshi et al., 2000).

$$\text{MFE} = \frac{\sum_{i=1}^{n} \chi_c(t) - \sum_{i=1}^{n} \chi_0(t)}{\eta} \quad (15)$$

where,

- $\chi_c(t)$ = Computed stream flow value
- $\chi_0(t)$ = Observed stream flow value
- $\eta$ = Number of observations

#### (b) Mean Absolute Error

$$\text{MAE} = \frac{\sum_{i=1}^{n} |\chi_c(t) - \chi_0(t)|}{\eta} \quad (16)$$

#### (c) Mean Relative Error

$$\text{MRE} = \frac{\sum_{i=1}^{n} |\chi_c(t) - \chi_0(t)|}{\chi_0(t)} \quad (17)$$

#### (d) Mean Square Error

$$\text{MSE} = \frac{\sum_{i=1}^{n} [\chi_c(t) - \chi_0(t)]^2}{\eta} \quad (18)$$

#### (e) Root Mean Square Error

$$\text{RMSE} = \left[\frac{\sum_{i=1}^{n} [\chi_c(t) - \chi_0(t)]^2}{\eta}\right]^{1/2} \quad (19)$$

#### (f) Integral Square Error
Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

\[
ISE = \sqrt{\frac{\sum_{t=1}^{n} (x_{i}(t) - \hat{x}_{0}(t))^2}{\sum_{t=1}^{n} \hat{x}_{0}(t)}}
\]  \hspace{1cm} (20)

D. Goodness of fit of autoregressive (AR) models

The following tests were performed to test the goodness of fit of autoregressive (AR) models.

(a) Box-Pierce Portmonteau lack of fit test

\[
Q = N \sum_{k=1}^{L} r_k^2
\]  \hspace{1cm} (21)

Where,

- \( N \) = Number of observations
- \( r_k \) = Serial correlation or autocorrelation of series \( Y_t \)

The statistic \( Q \) follows \( \chi^2 \) distribution with \( r = K-p \) degree of compared with tabulated values of \( \chi^2 \).

(b) Akaike Information Criterion

The Akaike Information Criterion (Akaike, 1974) was used for checking whether the order of the fitted model is adequate compared with the order of dependence model. Akaike Information Criterion for AR(p) models, was computed using the following equation.

\[
AIC (P) = N \ln \left( \frac{\sum \hat{\epsilon}^2}{\sigma^2} \right) + 2(P)
\]  \hspace{1cm} (22)

Where,

- \( N \) = Number of observations
- \( \hat{\epsilon}^2 \) = Residual variance
- \( \sigma^2 \) = Residual variance

A comparison was made between the AIC (p) and the AIC (p-1) and AIC (p+1). If the AIC (p) is less than both AIC (p-1) and AIC (p+1), then the AR (p) model is best otherwise, the model which gives minimum AIC value was the one to be selected model.

III. RESULT AND DISCUSSION

The standardized annual rainfall and runoff series were modeled through the autoregressive model. 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. (Salas and Smith 1980 b) Autocorrelation and partial autocorrelation are used for identification of the proper type and order of the model. The identification generally depends on the characteristics of overall water resources system, the characteristics of time series and the models input. (Salas and Smith 1981) demonstrated these of physical consideration of the type of 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 up to 4 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 1-2 ;figs 1-3) . The predicted values of annual rainfall and runoff were compared with the observed values. It was observed that AR(p) model up to 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
AR (1) : \( Y_t = 1050.81 + 0.7876(Y_{t-1} - 1050.81) + 0.773 \)

AR (2) : \( Y_t = 1050.81 + 0.8390(Y_{t-1} - 1050.81) + 0.773 \)

Table 1: Statistical parameters of autoregressive (AR) model for rainfall

<table>
<thead>
<tr>
<th>Model</th>
<th>AR(0)</th>
<th>AR (1)</th>
<th>AR (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoregressive parameter</td>
<td>-</td>
<td>( \Phi_1 = 0.7876 )</td>
<td>( \Phi_1 = 0.8390 ) ( \Phi_2 = 0.4546 )</td>
</tr>
<tr>
<td>White noise variance</td>
<td>99107.87</td>
<td>74675.89</td>
<td>70986.24</td>
</tr>
<tr>
<td>Akaike Information Criteria</td>
<td>69.02</td>
<td>69.33</td>
<td>71.02</td>
</tr>
<tr>
<td>Value of port monteau statistics</td>
<td>0.0294</td>
<td>1.5</td>
<td>0.4374</td>
</tr>
<tr>
<td>Degree of freedom upto 4 lags</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Table value of ( \chi^2 ) at 5% level of significance</td>
<td>9.48</td>
<td>7.815</td>
<td>5.991</td>
</tr>
</tbody>
</table>

Table 2: Statistical parameters of autoregressive (AR) model for runoff

<table>
<thead>
<tr>
<th>Model</th>
<th>AR(0)</th>
<th>AR (1)</th>
<th>AR (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autoregressive parameter</td>
<td>-</td>
<td>( \Phi_1 = -0.2302 )</td>
<td>( \Phi_1 = -0.2546 ) ( \Phi_2 = -0.0258 )</td>
</tr>
<tr>
<td>White noise variance</td>
<td>958.02</td>
<td>1089.95</td>
<td>913.27</td>
</tr>
<tr>
<td>Akaike Information Criteria</td>
<td>41.19</td>
<td>43.9</td>
<td>44.90</td>
</tr>
<tr>
<td>Value of port monteau statistics</td>
<td>2.923</td>
<td>1.3254</td>
<td>0.6613</td>
</tr>
<tr>
<td>Degree of freedom upto 4 lags</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Table value of ( \chi^2 ) at 5% level of significance</td>
<td>9.48</td>
<td>7.815</td>
<td>5.991</td>
</tr>
</tbody>
</table>
Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

Box Pierce Portmonteau and Akaike Information criterion test for AR models
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 result 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 acceptable 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 function against lag k. A graphical comparison of observed and predicted annual rainfall and runoff with the selected model are shown in fig 2 and 4. The graphical representation of the data shows a closer agreement between observed and predicted annual rainfall and runoff selected model. It reveals that developed model for rainfall and runoff can be utilized for the prediction of future trends with the minimum chance of error.

Fig 1 Comparison of correlogram of measured and predicted series for Rainfall

Fig 2 Comparison between measured and predicted annual rainfall of Kelowatershed

Box Pierce Portmonteau and Akaike Information criterion test for AR models
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 result 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 acceptable 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 function against lag k. A graphical comparison of observed and predicted annual rainfall and runoff with the selected model are shown in fig 2 and 4. The graphical representation of the data shows a closer agreement between observed and predicted annual rainfall and runoff selected model. It reveals that developed model for rainfall and runoff can be utilized for the prediction of future trends with the minimum chance of error.
Fig 3 Comparison of correlogram of measured and predicted series for runoff

Fig 4 Comparison between measured and predicted annual runoff of Kelo watershed

Statistical characteristics of data

The mean, standard deviation and skewness of historical and generated data was calculated to evaluate the fitting of the model in moment preservation. The results are tabulated in Table 3. The results clearly shows that the skewness of generated data by AR(1) model and historical data is lying between -1 to +1 and therefore AR(1) model preserved the mean and skewness better.

Table 3 Statistical characteristics of observed and predicted rainfall and runoff for Kelo watershed

<table>
<thead>
<tr>
<th>Statistical characteristics</th>
<th>Rainfall Observed</th>
<th>Rainfall Predicted</th>
<th>Runoff Observed</th>
<th>Runoff Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>991.29</td>
<td>983.33</td>
<td>163.83</td>
<td>151.425</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>397.31</td>
<td>387.216</td>
<td>44.87</td>
<td>26.687</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.699</td>
<td>0.870</td>
<td>-0.4535</td>
<td>-0.8562</td>
</tr>
<tr>
<td>Mean forecast error</td>
<td>-4.617</td>
<td>-</td>
<td>-</td>
<td>-6.294</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>4.617</td>
<td>0.070</td>
<td>-</td>
<td>6.294</td>
</tr>
<tr>
<td>Mean relative error</td>
<td>-</td>
<td>0.576</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Root mean square error</td>
<td>319.818</td>
<td>17.883</td>
<td>0.011</td>
</tr>
<tr>
<td>Integral square error</td>
<td>594.305</td>
<td>24.378</td>
<td>0.093</td>
</tr>
</tbody>
</table>

Performance Evaluation of AR(1) model for rainfall and runoff
Performance of the model was also estimated by estimating statistical characteristics such as MFE, MAE, MRE, MSE, RMSE and ISE were also used to prove the adequacy of the model for future prediction with higher degree of correlation to previous measured observations. In rainfall and runoff, the MFE for AR(1) model is -4.617 mm and -6.294 mm respectively. The different errors between measured and predicted values by AR (1) for both rainfall and runoff 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 Kelo watershed.

IV. CONCLUSION
On the basis of 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 Kelo watershed of Chhattisgarh state.

REFERENCE