Available online at www.ijpab.com

ISSN: 2320 – 7051
Int. J. Pure App. Biosci. SPI: **6 (1):** 171-178 (2018)

Research Article

International Journal of Pure & Applied Bioscience

OPEN ACCESS

Development of Autoregressive Time Series Model for Prediction of Runoff in Kachhinda Watershed Morena

S. S. Basediya[1*], Satish K. Singh[2], Mahendra Jadia[3], Akhilesh Solanki[4], C. J. Wesley[5] and Jitendra Bairwa[6]

[1,2,4]Department of Soil Water Engineering, JNKVV, Jabalpur, India

[1]Professor, Vaugh School of Agricultural Engineering and Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad, U.P., India

[2]Research Scholar, Department of Horticulture, MGCGV Chitrakoot Satna M.P. India

[3]Department of Farm Machinery and Power Engineering, JNKVV, Jabalpur, India

Received: 7.07.2018 | Revised: 16.08.2018 | Accepted: 27.08.2018

ABSTRACT

A study was conducted to develop a stochastic time series model, capable of prediction of runoff in Kachhinda watershed .The Kachhinda watershed of Sod river tributary of Chambal river situated in Gird region of Madhya Pradesh state. The Auto regressive time series model between observed and estimated runoff was developed and also observed that the correlation coefficient was found to be 0.974. Autoregressive model of order 0, 1 and 2 were tried for annual stream flow series and the annual stream flow was predicted. The goodness of fit and adequacy of model were tested by box- pierce portmanteau test, Akaike information criterion and by comparison of historical and generated data correlogram. For runoff the AIC value for AR (1) model is (0.91958) which lying between AR (0) is (0.207433) and AR (2) is (5.9767). The least which AR (1) is satisfying the selection criteria. The mean forecast error is also very less in case of runoff in AR (1) model on the bases of the statistical test, AIC the AR (1) model with estimated model parameters was estimated for the best future prediction in Kachhinda watershed.

Key words: *Autoregressive (AR) models, Surface runoff, Akaike information Criterion, Prediction of runoff and Box- Pierce Portmanteau test.*

INTRODUCTION

A watershed is a basin-like landform defined by highpoints and ridgelines that descend into lower elevations and stream valleys. A watershed is a basin like landform defined by peaks which are connected by ridges that descend into lower elevations and small valleys. A watershed carries water from the land after rain falls and snow melts. Water is channeled into soils, groundwater, creeks, and streams, making its way to larger rivers and eventually the sea. Rain water falling on it drop by drop and channels into soil, rivulets and streams flowing into large rivers and in due course sea.

Cite this article: Basediya, S.S., Singh, S.K., Jadia, M., Solanki, A., Wesley, C.J. and Bairwa, J., Development of Autoregressive Time Series Model for Prediction of Runoff in Kachhinda Watershed Morena, *Int. J. Pure App. Biosci.* SPI: **6(1):** 171-178 (2018).

Time series modeling is a dynamic research area which has attracted attentions of researchers community over last few decades. The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past[12]. Due to the indispensable importance of time series forecasting in numerous practical fields such as business, economics, finance, science and engineering, etc[15,16,13]. proper care should be taken to fit an adequate model to the underlying time series. It is obvious that a successful time series forecasting depends on an appropriate model fitting. A lot of efforts have been done by researchers over many years for the development of efficient models to improve the forecasting accuracy. As a result, various important time series forecasting models have been evolved in literature.

One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA)[15,16,7,6]. model. The basic assumption made to implement this model is that the considered time series is linear and follows a particular known statistical distribution, such as the normal distribution. ARIMA model has subclasses of other models, such as the Autoregressive (AR)[3,6,9], Moving Average (MA)[3,6] and Autoregressive Moving Average (ARMA)[3,7,6] models. For seasonal time series forecasting, Box and Jenkins[3] had proposed a quite successful variation of ARIMA model, viz. the Seasonal ARIMA (SARIMA)[5,3,6]. The popularity of the ARIMA model is mainly due to its flexibility to represent several varieties of time series with simplicity as well as the associated Box-Jenkins methodology[5,3,16,6]. for optimal model building process. But the severe limitation of these models is the pre-assumed linear form of the associated time series which becomes inadequate in many practical situations. To overcome this drawback, various non-linear stochastic models have been proposed in literature[15,16,10]. however from implementation point of view these are not so straight-forward and simple as the ARIMA models.

MATERIAL AND METHODS

The Kachhinda watershed is located between 25010’ East to 33015’ East longitude and 33027’ North to 37098’ North latitude. The total area of code is 600.00 hectare. The shape of watershed is nearly rectangular. The total area is 600.00 hectare out of which 59 hectare area is under agriculture use. and 499 hectare area is culturable waste land and 42 hectare area is non culturable wast land. The project area falls under the Central Chamble river Alluvial Plain of Madhya Pradesh, which is a level plain densely populated and most parts of the land is available for cultivation The valleys of the larger rivers are not only depressed well below the general level of the country but are of considerable breadth. Thus there is a wide area of low land which is inundated in years of the watershed lies in the sub-tropical climate. The average annual rainfall 720 mm. Most of the annual rain fall (about 80%) is received during the rainy season (July to September) accompanied with high intensity storm the heavy rainfall.

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.

$$Y_t = \overline{Y} + \sum_{j=1}^{p} \Phi_j (Y_{t-j} - \overline{Y}) + \varepsilon_t \qquad \text{………..(1)}$$

Where,

Yt = The time dependent series (variable)

ε_t = The time independent series which is independent of Yt and is normally distributed with mean zero and variance σ^2

Ȳ = Mean of normal rainfall and runoff data

$\Phi_1, \Phi_2 ....\Phi_p$ = Autoregressive parameter

B. Estimation of Autoregressive parameter (Φ) maximum likelihood estimate.

For estimation of the model parameter method of maximum likelihood will be used[3]

zizj+ zi+1zj+1+................ +zn+1-j zn+1-1 …..…………..(2)

and define

Dij= Dji= $\frac{N}{N+2-i-j}\sum_{l=0}^{N+1-(i+j)}$ zi+l zl+ I ……………..………..………..(3)

where,

D=difference operator

N=sample size

i,j=maximum possible order

AR (1) : $\Phi_1 = \frac{D_{1,2}}{D_{2,2}}$ ....(4)

AR (2) : $\Phi_1 = \frac{D_{1,2}D_{3,3} - D_{1,3}D_{2,3}}{D_{2,2}D_{3,3} - D_{2,3^2}}$ ……………..………..………..(5)

$\Phi_2 = \frac{D_{1,3}D_{2,2} - D_{1,2}D_{2,3}}{D_{2,2}D_{3,3} - D_{2,3^2}}$ ………..(6)

(a). Autocorrelation function

The autocorrelation function rk of the variable Yt of equation (3.2) is obtained by multiplying both sides of the equation (1) by Yt+k and taking expectation term by term. The relationship proposed by[8]. for the computation of autocorrelation function of lag K was used which is expressed as:

$$r_k = \frac{\sum_{t=1}^{N-K}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{N}(Y_t - \overline{Y})^2}$$ …………………………(7)

where,

rk=Autocorrelation function of time series Yt at lag k

Yt=Stream flow series (historical data)

$\overline{Y}$ =Mean of time series Yt

k=Lag of K time unit

N=Total number of discrete values of time series Yt

The following equation was used to determine the 95 percent probability levels[2].

$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-K-1}}{N-K}$ …..(8)

where, N=Sample size. And K = Number of lag.

(b). Partial Autocorrelation function

The following equation was used to calculate the partial autocorrelation function of lag [4].

$$PC_{k,k} = \frac{r_k - \sum_{j=1}^{k-1} PC_{k-1,j}r_{k-j}}{1 - \sum_{j=1}^{k-1} PC_{k-1,j}r_j}$$ ……..…(9)

where,

PCk,k=Partial auto correlation function

rk= Autocorrelation function of time series Yt at lag k

PCk,j= PCk-1,j- PCk,k PCk-1,k-j ………..…..……………..…(10)

The 95 percent probability limit for partial autocorrelation function was calculated using the following equation[2].

$$PC_{k,k}(95\%) = \frac{1.96}{\sqrt{N}}$$ …………………..…………(11)

(c). Parameter estimation of AR (p) models

The average of sequence Yt was computed by following equation:

$$\overline{Y} = \frac{1}{N}\sum_{t=1}^{N} Y_t$$ …..……………..…..………(12)

The average $\sigma^2{}_{\varepsilon} = \frac{1}{(N-1)}\sum_{t=1}^{N}(Y_t - \overline{Y})^2$ ………………......………(13)

After computation of $\overline{Y}$ and $\sigma^2{}_{\varepsilon}$, the remaining parameters $\Phi_1, \Phi_2, \ldots\ldots\ldots\Phi_P$ of the AR models were estimated by using the sequence:

$Z_t = Y_{t-}\overline{Y},$

The parameters $\Phi_1, \Phi_2, \ldots\ldots\ldots\Phi_P$ were estimated by solving the p system of following linear equations [14].: $r_k = \Phi_1 r_{k-1} + \Phi_2 r_{k-2} + \ldots\ldots\ldots\ldots \Phi_P r_{k-P}$

$r_k = \sum_{j=1}^{P} \Phi_j r_{k-j}$ K>0 ……(16)

Where, r1, r2 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, runoff. The mean forecast error (MFE) was computed for the annual rainfall and runoff by the following equation[11].

$$\text{MFE} = \frac{\sum_{i=1}^{n}\chi_c(t) - \sum_{i=1}^{n}\chi_0(t)}{\eta} \qquad \ldots\ldots\ldots\ldots \quad (17)$$

where,

χ_c (t) = Computed rainfall and runoff value

χ_0 (t) = Observed rainfall and runoff value

η = Number of observations

(b) Mean Absolute Error

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|\chi_c(t) - \chi_0(t)\right|}{\eta} \qquad \ldots\ldots\ldots(18)$$

(c) Mean Relative Error

$$\text{MRE} = \frac{\sum_{i=1}^{n}\dfrac{\left|\chi_c(t) - \chi_0(t)\right|}{\chi_0(t)}}{\eta} \qquad \ldots\ldots\ldots\ldots(19)$$

(d) Mean square error

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left[\chi_c(t) - \chi_0(t)\right]^2}{\eta} \qquad \ldots\ldots\ldots(20)$$

(e) Root Mean Square Error

The root mean square error was computed by following equation[11].

$$\text{RMSE} = \left[\frac{\sum_{i=1}^{n}\left[\chi_c(t) - \chi_0(t)\right]^2}{\eta}\right]^{1/2} \qquad \ldots\ldots\ldots(21)$$

(f) Integral Square Error

$$\text{ISE} = \frac{\sqrt{\sum_{i=1}^{n}\left[\chi_c(t) - \chi_0(t)\right]^2}}{\sum_{i=1}^{n}\chi_0(t)} \qquad \ldots\ldots (22)$$

D. Goodness of fit of autoregressive (AR) models

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

E. Box-Pierce Portmanteau lack of fit test

The Box-Pierce Portmanteau lack of fit test was used to check whether the residual of a dependence model for correlation. The test statistic was computed by using the following equation:

$$Q = N \sum_{k=1}^{L} r_k^{\,2} \qquad \ldots\ldots\ldots(23)$$

Where,

N = Number of observations

Rk = Serial correlation or autocorrelation of series Yt

The statistic Q follows $\chi 2$ distribution with r = K-p degree of freedom. The estimated value of $\chi 2$ was compared with tabulated values of $\chi 2$.

F. Akaike Information Criterion

A mathematical formulation which considers the principle of parsimony in model building is the Akaike Information Criterion[1]. 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.

$$\text{AIC (P)} = \text{N 1n}\left(\hat{\sigma}_{\varepsilon}^{2}\right) + 2(P) \qquad \ldots\ldots\ldots(24)$$

Where,

N= Number of observation

$\hat{\sigma}_{\varepsilon}^{2}$ = 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.

RESULT AND DISCUSSION

The standardize annual rainfall and runoff series were modeled through the autoregressive 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 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 autoregressive models up to order 2 were tried in this study. The parameters of AR models up to order2 were determined through equation (1) and model is given as under:

Runoff:

AR (1): $Y_t = 119.06 + 0.9191858(Y_{t-1} - 119.06) + \varepsilon_t$

Table 1 represents autocorrelation of observed annual stream flow for runoff volume for Kachhinda watershed. The autocorrelation function is a graphical relationship of autocorrelation function rk (y) with lag k. The autocorrelation was used for identifying the order of the model for given time series. The autocorrelation coefficients at different lag were computed by equation 3.7 and SPSS Software. The 95 per cent probability limits of their coefficients were computed by equation 3.9. All values of Autocorrelation is lies between -1 to +1 so the result revealed that the existing pattern of autocorrelation further indicate the possibility to use AR (1) model.

Table 1. Observed and estimated runoff of the study area

Year	Observed Runoff (mm)	AR(estimated) Runoff (mm)
2001	112.8	115.42
2002	146.9	151.31
2003	76.2	80.37
2004	115.4	116.6
2005	153.3	157.27
2006	113.2	113.91
2007	140.1	141.14
2008	96.7	98.147
2009	125.4	130.51
2010	110.6	114.07

Table 2 shows the value of partial autocorrelation of observed annual stream flow for runoff volume for Kachhinda watershed. Partial autocorrelation is the plot of partial autocorrelation coefficients against lag k. To construct the partial autocorrelation, partial autocorrelation coefficient at different lag were computed using equation 3.9 or SPSS software. The 95 per cent probability limits of these functions were determined by equation 3.9. The value of partial autocorrelation function is vary from -1 to +1 so the result revealed that the autoregressive model of first order, AR (1) model was selected for further analysis.

Table 2. Autocorrelation of Observed Annual Stream Flows for Runoff Volume for Kachhinda Watershed.

Lag	95% Lower limit	Autocorrelation function	95% Upper limit
1	-0.727	0.727	0.504
2	-0.773	0.773	0.523
3	0.828	0.828	0.542
4	-0.897	-0.897	0.563
5	-0.984	-0.984	0.584

The statistical parameters of autoregressive (AR) models for runoff are given in table 3. In order to choose the better model among these three models Akaike Information Criterion

(AIC) for all three models were computed. If AIC (p) is less than both AIC (p-1) and (p+1) the AR (p) model is best, otherwise the model which gives the minimum AIC value was the one to be selected. Thus from the table it is clear that AIC value of AR (1) is lying between AR (2) and AR (0), therefore it is considered suitable model for the prediction of runoff.

The Box-Pierce Portmanteau lack of test fit was used to check the adequacy of autoregressive models for runoff. The values of statistical test for AR (0), AR (1) and AR (2) models. The test statistics were compared and it reveals that the value of test statistics for all three models viz, AR (0), AR (1) and AR (2) were giving good fit and were acceptable.

Table 3. Partial Autocorrelation of Observed Annual Stream Flows for Runoff Volume for Kachhinda Watershed

Lag	95% Lower	Partial Autocorrelation	95% Upper limit
1	-0.619	0.727	0.619
2	-0.619	0.704	0.619
3	-0.619	0.129	0.619
4	-0.619	-0.103	0.619
5	-0.619	-0.098	0.619

Table 4 and 5 represents statistical characteristics of observed and predicted annual stream flow for runoff and evaluation of regeneration performance with statistical errors respectively. The results clearly shows that the skewness of the generated data by AR (1) and historical data is lying between -1 to +1 and therefore AR (1) model preserved the mean and skewness better.

Table 4. Statistical Parameters of Autoregressive (AR) Models for Runoff

Modle	AR (0)	AR(1)	AR (2)
Autoregressive Parameters	-	Φ_1=0.91918	Φ_1= 5.9767 Φ_2 = 0.207433
White Noise Variance,	85.94935	2.6283	3.5862
Akaike Information criterion, AIC (P)	151.42	136.750	68.6987
Value of Porte Moniteau statistics, Q	91.9493	75.6283	16.5862
Degree of freedom upto 5 lags	5	4	3

Table 5. Statistical Characteristics of Observed and Estimated Annual Runoff

S. No.	Statistical Characteristics	Observed Runoff	Predicted Runoff
1	Mean	116.863	121.45
2	Standard Deviation	23.39459	23.39459
3	Swekness	-0.21395	-1.25703

The statistical characteristics such as MFE, MAE, MRE, MSE, RMSE and ISE were also used to test the adequacy of the model for future prediction with higher degree of correlation to previous measured observations. The different types of errors in runoff generation of AR (1) model are calculated and presented in Table 6. Although the error is slightly high but the model can be used for rainfall prediction in Kachhinda watershed. The value of MAE for runoff is equal to the respective MFE. This is due to reason that the difference in predicted and historical values is only one side in most of the years (predicted side).The data of the table clearly represents that prediction AR (1) model is giving the best results. It indicates that AR (1) model can used for rainfall and runoff prediction of the watershed.

Table 6. Evaluation of Performance with Statistical Errors

S. No	Statistical error	Autoregressive AR (1) model Runoff
1	Mean forest error	2040.76
2	Mean absolute error	-2040.76
3	Mean Relative error	-8.2938
4	Mean Square error	93937.965
5	Root Mean square error	306.493
6	Integral Square error	0.81405

Comparison of the Observed and Predicted Model Correlogram:

A graphical comparison of historical and predicted correlogram with the selected model is shown in Fig. 1. The graphical representation of the data shows a closer agreement between observed correlogram of runoff and predicted values by selected model while in case of runoff the historical correlogram is not showing so much close agreement at some points of the series. It also reveals that the developed model for runoff can be utilized for the prediction of future trends with minimum chance of error.

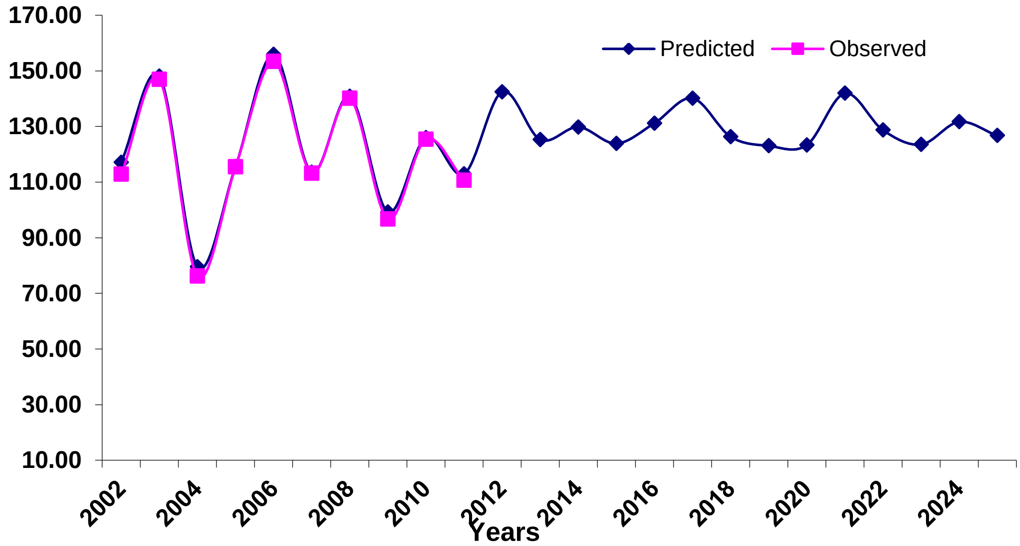

Fig. 1: Comparison of Correlograms of Observed and Predicted Runoff

CONCLUSION

The goodness of fit and adequacy of models were tested by Box-Pierce Portmanteau test, Akaike Information Criterion (AIC) and by comparison of historical and predicted correlogram. The AIC value for AR (1) model (136.750) is lying between AR (0) (151.427) and AR (2) (68.598) which is satisfying the selection criteria. The mean forecast error is also very less. On the basis of the statistical test, Akaike Information Criterion, AIC the AR (1) model with estimate model parameters was estimated for the best future predictions in Kachhinda watershed. This is also proved by Graphical representation between historical and generated correlogram, where in runoff there is a very close agreement.

REFERENCES

1. Akaike, H., "A new look at the statistical model identification". IEEE Transactions on automatic control, AS-19, pp. 716-723 (1974).
2. Anderson, R. L., "Distribution of the serial correlation coefficients". *Annals of*
3. *Math. Statistics.***13(1):** 1-13 (1942).
4. Box, G.E.P. and Jenkins, G.,"Time series Analysis, Forecasting and control".
5. First Ed. Holden – Day Inc., *San Francisco*. p 575 (1970).
6. Durbin, J., "The fitting of time series model". Rev. *Int. Inst. stat*., **28:** 23-33 (1960).
7. Hamzacebi, C.,"Improving artificial neural networks' performance in seasonal time series forecasting", *Information Sciences* 178, pp 4550-4559 (2008).
8. Hipel, K.W. and McLeod, A.I.,"Time Series Modelling of Water Resources and Environmental Systems", *Amsterdam, Elsevier*. 1013pp (1994).
9. John, H., Cochrane,"Time Series for Macroeconomics and Finance", Graduate School of Business, University of Chicago, spring. 134pp (1997).
10. Kottegoda, N. T. and Horder, M. A.," Daily flow model rainfall occuerces using pulse and a transfer function". *J. Hydrology*,**47:** 215-234 (1980).
11. Lee, J.,"Univariate time series modeling and forecasting (Box-Jenkins Method)", Pp 26-40 (1999).
12. Parrelli, R., "Introduction to ARCH & GARCH models", Optional TA Handouts, Econ 472 Department of Economics, University of Illinois, pp 1-7 (2001).
13. Raghuwanshi, S, and Wallender, W.S., "Forecasting Daily Evapotranspiration for a Grass Reference Crop". *Agricultural Engg. Journal*, **9(1):** 1-16 (2000).
14. Raicharoen ,T., Lursinsap C. and Sanguanbhoki, P.,"Application of critical support vector machine to time series prediction", *International Symposium on pp* 741-744 (2003).
15. Tong, H., "Threshold Models in Non-Linear Time Series Analysis", Springer-Verlag, New York, pp 1-38 (1983).
16. Yule, G.V.,".On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot number" .*Phil. Tran.* **226:** 267 (1927).
17. Zhang, G.P., "A neural network ensemble method with jittered training data for time series forecasting", *Information Sciences* 177, pp 5329–5346 (2007).
18. Zhang, G.P.,"Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing 50, pp159–175 (2003).