Int.J.Curr.Microbiol.App.Sci (2018) *7*(4): 1516-1522

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 7 Number 04 (2018)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2018.704.170

Prediction of Rainfall of Allahabad District by the Development of Autoregressive Time Series Model

K.N. Singh[1*], A. Dalai[1], M.P. Kaiwart[1], R.R. Mohanty[2] and B. Kumar[1]

[1]Faculty of SWE, SVCAET, IGKV, Raipur-492012, Chhatishgarh, India
[2](Agril. Engg.) of Agro Polytechnic Centre, Rourkela, OUAT, Bhubaneswar-751003, Odisha, India

**Corresponding author*

A B S T R A C T

The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic differential equation. The present study was conducted with the main objective to develop a stochastic time series model for prediction of rainfall of Allahabad district, which lies between $25^0$47' N latitude, $81^0$21'E longitude and elevation of 104 m from the mean sea level. The Geographical area of Allahabad district is 5246 km^2 and to determine the annual rainfall, rainfall data of 28 years from 1983 to 2010 were used to develop the Autoregressive (AR) time series models of orders 0, 1 and 2. The general recursive formula was used to determine various parameters of the model. The goodness of fit and adequacy of models were tested by Box-Pierce Portmonteau test, Akaike Information Criterion, by comparing observed and predicted correlogram. The AIC value for AR (1) model is lying between AR (0) and AR (2) which is satisfying the selection criteria. The close agreement in rainfall is observed from the graphical representation between observed and generated correlogram. The developed model can be used efficiently for the prediction of rainfall of Allahabad district, as observed from the comparison between the observed and predicted rainfall by AR (1) model. This method can be immensely helpful for the farmers and research workers for water harvesting, ground water recharge, flood control and development of water management strategies.

Keywords

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

Article Info

Accepted:
12 March 2018
Available Online:
10 April 2018

Introduction

Fresh water is an essential resource and rainfall is the primary source of fresh water supply. The process of precipitation is a fundamental component of the hydrological cycle and it is a complex and delicately balanced process (Shaw, 1988). Rainfall is a natural phenomenon resulting from atmospheric and oceanic circulation (local convection, frontal patterns) (Meher and Jha 2013).

A time series is defined as a set of observation arranged chronologically that is a sequence of observation usually ordered in time but may be ordered according to some other dimension. The principal aim is of time series analysis to describe the history of moments in time of some variable at a particular site. Most hydrologic system have both deterministic as well as stochastic component, but stochastic time series model such as autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) (Carlson,1970). In addition to this it may be noted that models found applicable in a particular zone *e.g.*, the temperate zone not to be applicable in other zone such as tropic (Iyengar, 1982). The present study is to develop a stochastic model of annual rainfall time series applicable for Allahabad District and has been planned with following objectives:

To estimate parameters of autoregressive model for Allahabad district.

To develop stochastic time series model for prediction of rainfall for Allahabad district.

To test the validity of the predicted rainfall with measured and to evaluate the performance of the developed model.

Materials and Methods

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: θ) with the parameters $\theta = (\theta_1, \theta_2 \ldots)$ valid for all positions t = 1, 2… N and without any dependence between $X_1, X_2 \ldots X_n$.

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

$\varepsilon_t = \phi\, \varepsilon_{t-1} + \eta_t \ldots\ldots$ (1)

Where,

η_t = An independent series with mean zero and variance $(1-\phi^2)$

ε_t= Dependent series

ϕ = Parameter of the model

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^{th} order autoregressive model AR (p), representing the variable Y_t is generally written as.

$$Y_t = \overline{Y} \sum_{j=1}^{p} \phi_j (Y_{t-j} - \overline{Y}) + \varepsilon_t \quad (2)$$

Where,

Y_t = The time dependent series (variable)

ε_t= The time dependent series which is independent of Y_t and is normally distributed with mean zero and variance σ_ε^2

Y = Mean of annual flow and rainfall data

$\Phi_1, \phi_2 \ldots \phi_p$= Autoregressive parameter

Estimation of Autoregressive parameter (ϕ) maximum likelihood estimate

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

Zi×zj + zi+1zj+1 +………+ zN+1-J zN+1-Iand define

$$Dij = Dji = \frac{N}{(N+2-i-j)} \quad (3)$$

Where

D = difference operator

N = sample size
i, j = maximum possible order

$$AR(1): \phi_1 = \frac{D1,2}{D2,2} \quad (4)$$

$$AR(2): \phi_1 = \frac{D1,2\ D3,3 - D1,3\ D2,3}{D2,2\ D3,3 - D^2\ 2,3} \quad (5)$$

$$AR(2): \phi_2 = \frac{D1,3\ D2,2 - D1,2\ D2,3}{D2,2\ D3,3 - D^2\ 2,3} \quad (6)$$

Autocorrelation function

The autocorrelation function r_k of the variable Y_t of equation (2) is obtained by multiplying both sides of the equation (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.

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

Where,

r_k= Autocorrelation function of time series Y_t at lag k

Y_t= Stream flow series (historical data)

$\bar{Y}$ = Mean of time series Y_t

k = Lag of K time unit

N = Total number of discrete values of time series Y_t

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

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \quad (8)$$

Where, N = Sample size

Partial autocorrelation function

The following equation was used to calculate the partial autocorrelation function of lag k

$$PC_{k,k} = \frac{r_k - \sum PC_{k-1,k-j}}{1 - \sum PC_{k-1,j}} \quad (9)$$

Where,

$PC_{k,k}$ = Partial autocorrelation function at lag K
r_k = Autocorrelation function at lag K
$PC_{k,j} = PC_{k-1,j} - PC_{k,k}.PC_{k-1,k-j}$ (10)
j = 1, 2 …k-1

The 95 percent probability limit for partial autocorrelation function was calculated using the following equation (Anderson 1942)

$$PCk,k(95\%) = \frac{1.96}{\sqrt{N}} \ldots\ldots (11)$$

Parameter estimation of AR (p) models

The average of sequence Y_t was computed by following equation

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

The average σ^2_ε of Y_t was computed by the following equation

$$\sigma_\varepsilon^2 = \frac{1}{(N-1)} \sum_{t-1}^{N} (Y_t - \bar{Y})^2 \quad (13)$$

After computation of Y and σ_ε^2 the remaining parameters $\phi_1, \phi_2, \ldots\ldots\phi_p$ of the AR models were estimated by using the sequence.

$Z_t = Y_t - Y$ (14)

$t = 1, 2, \ldots..n$

The parameters $\phi_1, \phi_2, \ldots\ldots\phi_p$ were estimated by solving the p system of following linear equation (Yule and Walker equation).

$r_k = \phi_1 r_{k-1} + \phi_2 r_{k-2} + \ldots\ldots + \phi_p r_{k-p}$ k>0

$$r_k = \sum_{j=1}^{p} \emptyset_j r_k - 1 \quad \ldots\ldots(15)$$

Where, r_1, r_2 were computed from equation (7)

Goodness of fit of Autoregressive (AR) models

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

Box-Piece Portmanteau lack of fit test

$$Q = N \sum_{k=1}^{L} rk^2 \quad (22)$$

Where,

N = Number of observation

r_k = serial correlation or Autocorrelation of series Y_t

The statistic Q follows χ^2 distribution with r = k-p degree of freedom.

The estimated value of χ^2

Akaike information criterion

Akaike Information Criterion for AR (p) models was computed using the following equation.

$AIC (P) = N \ln(\sigma_E^2) + 2 (p)$ (23)

Where,

N = Number of observations

σ_E^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.

Results and Discussion

The study annual rainfall series were modelled through the autoregressive model. The modelling 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 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 5 lag of the series (lag k) were computed and the autoregressive model of first order AR(1) was selected for further analysis.

Fig.1 Comparison of correlogram of observed, normalized and predicted series for rainfall

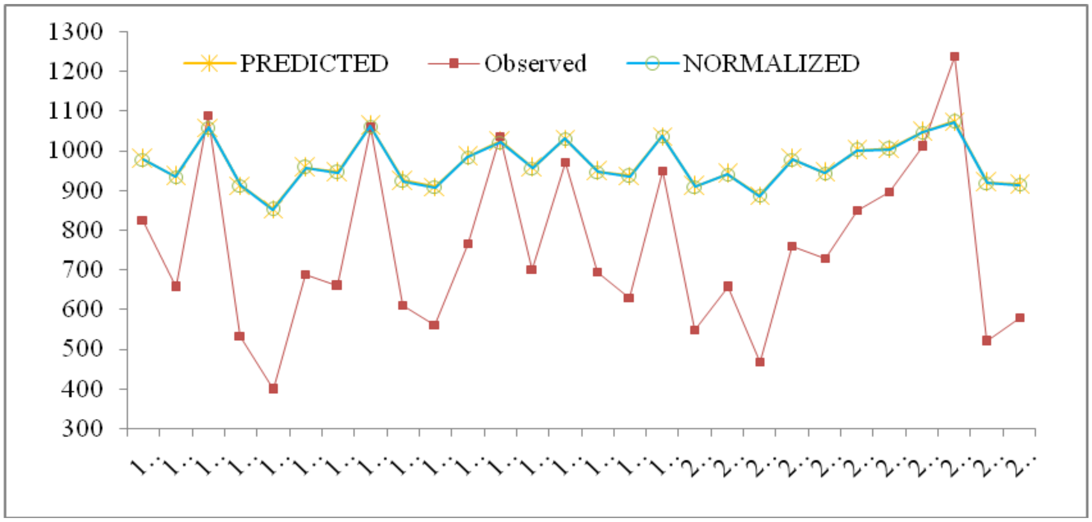

Fig.2 Comparison between observed and predicted annual rainfall for Allahabad district

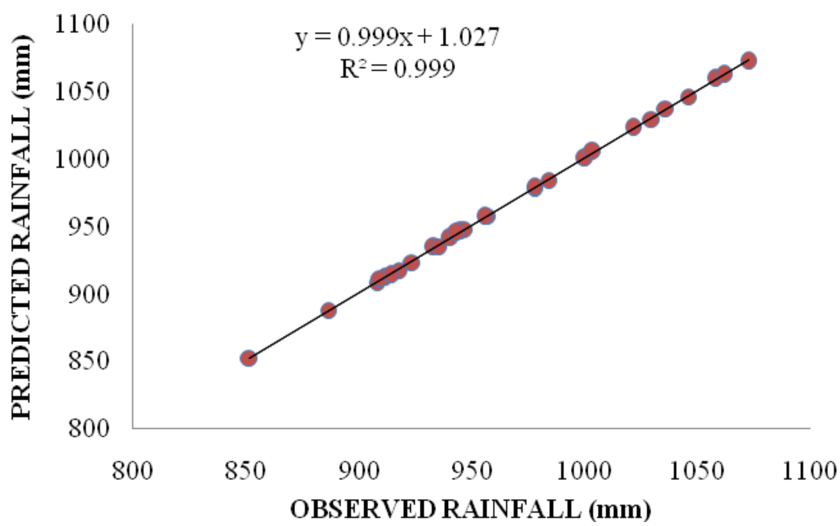

Table.1 Statistical parameters of autoregressive (AR) model for rainfall

Model	**AR(0)**	**AR (1)**	**AR (2)**
Autoregressive parameter	-	Φ_1= -0.0057	Φ_1 = -0.00167 Φ_2 =-0.13948
White noise variance	3340.934	3156.154	3221.726
Akaike Information Criteria	229.1922	229.5991	226.1748
Value of port monteaue statistics	1.05435	0.98685	0.872775
Degree of freedom upto 5 lags	5	4	3
Table value of χ^2 at 5% level of significance	11.07	9.48	7.81

Table.2 Statistical characteristic of observed and predicted rainfall

Sl. No.	Statistical characteristic	Observed rainfall, mm	Predicted rainfall, mm
1.	Mean	57.84722	57.7207
2.	Standard deviation	964.3575	965.3553
3.	Skewness	0.259169	0.7246

Table.3 Evaluation of regeneration performance with statistical errors

Sl. No.	Statistical error	Autoregressive (AR 1) model Rainfall (mm)
1	Mean Forecast Error	0.0353
2	Mean Absolute Error	-0.00148
3	Mean Relative Error	-0.0353
4	Mean Square Error	0.03489
5	Root mean square Error	0.186789
6	Integral Square Error	0.04721

Models of Autoregressive (AR) family

The parameters of AR model were computed for annual rainfall and the predicted values of annual rainfall 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 figures 1 and 2 AR (p) models for prediction of annual rainfall.

AR (1): $Yt = 966.0165 - 0.0057 (Y_{t-1} - 966.0165) + \varepsilon_t$

AR (2): $Yt = 966.0165 - 0.00167 (Y_{t-1} - 966.0165) + \varepsilon_t$

Box Pierce Portmonteau test for AR model

The Box–Pierce Portmonteau lack of fit test was used to check the adequacy of autoregressive models for both river flow and rainfall. The values of statistical tests for AR (0), AR (1) and AR (2) models were estimated. All the 3 models viz. AR (0), AR (1) and AR (2) were giving fit and were acceptable.

Comparison of the observed and predicted annual rainfall

The correlogram of observed and predicted series for annual rainfall were developed by plotting autocorrelation function against lag k. A graphical comparison of observed and predicted annual rainfall with the selected model is shown in figure 1. The graphical representation of the data in between observed and predicted annual rainfall selected model.

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 in Table 2. The results clearly shows that the skewness of generated data by AR (0) model and historical data is lying between -1 to +1 and therefore

AR (1) model preserved the mean and skewness better.

Performance evaluation of AR (1) model for rainfall

The performance of the model was estimated by estimating statistical characteristics such as MFE, MAE, MRE, MSE, RMSE and ISE were also estimated to prove the adequacy of the model for future prediction with higher degree of correlation to previous measured observations. In case of rainfall, the MFE for AR (1) model is 0.0353 mm which is lower the error, higher the quality of predicted rainfall. As the error is minimum, so the AR (1) model can be best suited for rainfall prediction for Allahabad district.

On the basis of estimated error, statistical characteristics and correlation between observed and predicted values, Autoregressive AR (1) model can be used to predict annual rainfall for Allahabad district.

References

Akaike, H., (1974). A new look at the statistical model identification on IEEE transaction on automatic control, AS-19. pp. 716-723.

Anderson, R. L., (1942). Distribution of the serial correlation coefficients. Annals of Math. Statistics. 13 (1):1-13.

Box, G.E. P and Jenkins, G. M., (1970). Time series Analysis Forecasting and control. First Ed. Holden- day Inc., San Francisco. Pp. 575.

Carlson, R.F (1970). Application of linear models to four annual stream flow series. Water Resource Research, 64:1070-1078

Iyengar, R. N., (1982). Stochastic modeling of monthly rainfall. Journal of Hydrology. Vol. 57. Pp. 375-387.

Kottegoda and Horder (1980). Daily flow model rainfall occurs using pulse and transfer function. J. Hydrology, 47: 215-234.

Meher and Jha (2013). SARIMA models for temperature and precipitation time series in the Czech Republic for the period 1961–2008. Journal of Applied Mathematics, 4(3): 281–290.

Salas, J.D. and Smith, R. A., (1980). Physical basis of stochastic models of annual flows. Water Resources Research, 17 (2): 428-430.

Shaw, E. M., (1988). Hydrology in Practice. Chapman & Hall, Second edition.

How to cite this article:

Singh, K.N., A. Dalai, M.P. Kaiwart, R.R. Mohanty and Kumar, B. 2018. Prediction of Rainfall of Allahabad District by the Development of Autoregressive Time Series Model. *Int.J.Curr.Microbiol.App.Sci.* 7(04): 1516-1522. doi: https://doi.org/10.20546/ijcmas.2018.704.170