Vol. 46, No. 2, pp 181-189, 2018

Indian Journal of Soil Conservation

Online URL : http://indianjournals.com/ijor.aspx?target=ijor:ijsc&type=home

Modelling monthly rainfall time series for Rahuri region in Maharashtra, India

Anil Kumar

Department of Soil and Water Conservation Engineering, G.B. Pant University of Agriculture and Technology, Pantnagar, Uttarakhand.

E-mail: anilkumar_swce61@rediffmail.com

ARTICLE INFO

Article history:

Received : May, 2017

Revised : May, 2018

Accepted : June, 2018

Key words:

Harmonic analysis,

Monthly rainfall time series,

Stochastic component,

Trend analysis

ABSTRACT

A stochastic model of monthly rainfall time series based on rainfall data for the years 1975 to 2010 was developed for Rahuri region in Maharashtra, India. The results indicate no deterministic trend component at 95% confidence level. However, the periodic component was found to be of 4 significant harmonics on the basis of Fourier series analysis. The stochastic components in the series were determined by autoregressive (AR) model of order six using the Statistical Application System (SAS-9.2) software. The independent stochastic component was found to be left skewed and the residual series was transformed by Box-Cox transformation into the normalized form. The developed time series model was used for regeneration and short-term forecast of monthly rainfall for 2009 and 2010. Results indicated that the properties propagated in the regenerated and predicted rainfall data series were comparable to those of the historical series which was revealed by integral square error (ISE), weighted pearson moment (WPM) and coefficient of efficiency (CE) as 3.53, 0.76 and 0.97 in calibration period and 3.05, 0.94 and 0.99 in validation period, respectively.

1. INTRODUCTION

Planning and designing of water resources projects needs accurate information of different hydrologic events, which are not governed by the known physical and chemical laws. Many hydrologic phenomena are highly erratic, complex and random in nature and hence they can be interpreted only in a probabilistic sense. One of the important problems in hydrology is to deal with interpretation of past record of hydrologic events in terms of future probabilities of occurrence. Hydrologic modelling of meteorological data as a stochastic process is of interest for a variety of hydrological research areas such as flood routing, reservoir operation and agricultural planning at watershed scale. The most common type of climatic data required by hydrologic model is the historical records of rainfall. Long sequences of rainfall data are essential components for better decision making on proper planning and management of soil and water conservation (SWC) projects to avoid the risk of either under-designing or uneconomical cost of over-designing at particular sites. Therefore, as an alternative, the rainfall records can be simulated using stochastic rainfall models (Haan *et al*., 1976; Panigrahi, 1998; Khatua *et al.*, 2014).

Most of the hydrologic systems have both deterministic and stochastic components. Stochastic model can be developed to generate alternative hydrologic data sequences that are likely to occur in future based on the characteristics of past historical records. Time series analysis and forecasting has become a major tool in various applications in hydrology and environmental management fields by making mathematical models to generate synthetic hydrologic records, to forecast hydrologic events, to detect trends and shifts in hydrologic records and to fill in missing data and extend records. Time series analysis is of immense practical use in dealing with forecasting of various hydrological parameters (Louck *et al.*, 1981).

Kottegoda and Horder (1980) proposed a probability model for daily rainfall by considering occurrences of wet and dry days and a rainfall-runoff model was developed using pulses and transfer function. Sapolia *et al.* (1980) suggested a simple method to determine parameters of an AR model with a real and equal root based on serial correlation coefficients. Several studies have been conducted using different approaches such as Fourier analysis, deterministic mathematical model, Markov-chain exponential model, autoregressive moving average (ARMA) model, principle of maximum entropy and autoregressive integrated moving average (ARIMA) model (Sen, 1980; Richardson, 1981; Haltiner and Salas, 1988; Krastanovic and Singh 1991; Gupta and Singh, 2004; Raghuwanshi and Wallender, 2000; Jayawardena *et al*., 2006; Kumar, *et al*., 2007).

Dabral *et al.* (2008) developed a monthly stochastic model using rainfall data of eighteen years for a station in northeastern state of Arunachal Pradesh, India and found the series to have three harmonics in the periodic component. Tularam and Ilahee (2010) examined long term data from coastal areas of Queensland. The study examined whether a cyclic nature existed in the data sets and whether a simple trend relationship between rainfall and temperature existed. A large data set involving more than 50 years of rainfall and temperature data were examined using spectral analysis, time series analysis-ARIMA methodology to analyse climatic trends and interactions. There is a cyclic pattern noted in both the rainfall and temperature time series.

Based on the above researches, it seems important to analyze the time series of various types of hydrological data so that forecasting of future events could be ascertained. A statistical software tool, SAS-9.2 was used for the analysis to avoid complicated and long manual computations. An attempt has been made in this study to develop a stochastic model of monthly rainfall time series to generate sequences of monthly rainfall data using the developed model and to compare its characteristics with the historical data; and to make short term forecasting of future events of rainfall for Rahuri region in Maharashtra, India.

2. MATERIALS AND METHODS

Collection and Analysis of Data

The daily rainfall data for Rahuri raingauge station for a period of 36 years (1975-2010) were collected from Meteorological Department, Mahatma Phule Krishi Vidyapeeth (MPKV), Rahuri, Maharashtra, India located 511 m above mean sea level at (msl) 74°39'0"E longitude and 19°23'0"N latitude. Monthly rainfall data for the period 1975 to 2008 were considered for the development of the mathematical model, and monthly rainfall data for the period 2009 to 2010 were used for the prediction of the model.

Analysis of continuously recorded time series is performed by transforming the continuous series into discrete time with a fixed interval. The monthly time series of rainfall are mostly non-stationary. To smoothen the recorded rainfall series Z_t, the square root transformation as suggested by Richardson (1978) was used as:

$$X_t = \sqrt{Z_t} \quad ...(1)$$

Where, X_t is transformed monthly rainfall series at time t; and Z_t is observed monthly rainfall series at time t.

A time series can be divided into the deterministic components consisting of a non-periodic (trend) and periodic or cyclic components, and the stochastic components consisting of chance dependent and independent components. Trend is the steady and regular movement in a time series through which the values, on an average, are either increasing or decreasing; whereas the periodic effects are short-term fluctuations which occur in a series due to imposition of a cyclic phenomenon and are repetitive over a fixed period of time. To check the presence of trend, the Turning point test and Mann Kendall's rank correlation test were conducted on the monthly data series at 5% level of significance (Patra, 2008; Panigrahi and Panigrahi, 2013). If the trend component is present, it should be separated from the raw data series for further analysis. The periodic component was determined for the residual trend-free series.

The periodic component was determined by the harmonic analysis, which required identification of base period and computation of significant number of harmonics. The identification of the base period was done by correlogram analysis. Each component of time-series model was systematically identified and removed to develop the final mathematical model of the time series.

Correlogram Analysis

A serial correlogram is a graphical relationship of autocorrelation function r_k with lag K. It measures the degree of linear dependence between the events k time units apart within the time series. The auto-correlation function r_k of series Y_t were computed using the relationship given by Kottegoda (1980):

$$r_K = \frac{\left[\sum_{t=1}^{N-K}\left\{\left(Y_t - \overline{Y}\right)\left(Y_{t+K} - \overline{Y}\right)\right\}\right]}{\left(\sum_{t=1}^{N}\left(Y_t - \overline{Y}\right)^2\right)} \quad ...(2)$$

Where, r_k is auto-correlation function of time series Y_t at lag K; Y_t is the rainfall series after removing trend component; $\bar{Y}$ is mean of time series Y_t; K is lag of time unit; and N is total number of discrete values of time series Y_t. The base period in the series was obtained by visual inspection of peaks and troughs in the auto-correlogram.

Harmonic Analysis

The periodic component in the mean from the set of observations for $Y_1, Y_2, Y_3, \ldots\ldots Y_N$ was estimated by the equation (Mutreja, 1986):

$$\mu_t = m_y + \sum_{j=1}^{m}\left(A_j \operatorname{Cos}\frac{2\pi\ j\tau}{\omega} + B_j \operatorname{Sin}\frac{2\pi j\tau}{\omega}\right) \quad ...(3)$$

Where, μ_τ is the periodic component in mean monthly value; m_y is the mean of series Y_t with t = 1, 2,…., N (number of discrete values of rainfall); A_j and B_j are Fourier coefficients of mean series; τ = 1, 2, 3,….., ω, the base period (= 12 for monthly rainfall data); j = 1, 2, 3, ….., m, the number of significant harmonics. The parameters A_j and B_j in eq. (3) were computed by the equations:

$$A_j = \frac{2}{\omega}\sum_{\tau=1}^{\omega} m_\tau \operatorname{Cos}\left(\frac{2\pi j\tau}{\omega}\right) \quad ...(4)$$

$$B_j = \frac{2}{\omega}\sum_{\tau=1}^{\omega} m_\tau \operatorname{Sin}\left(\frac{2\pi j\tau}{\omega}\right) \quad ...(5)$$

Where, m_τ represents monthly mean values for n number of years of record, computed as:

$$m_\tau = \frac{1}{n}\sum_{p=1}^{n} Y_{p,\tau} \quad ...(6)$$

Where, $Y_{p,\tau}$ is monthly rainfall in year p at position τ in the year.

Similarly, the periodic component in the standard deviation was also calculated by using the relationship:

$$\sigma_\tau = S_y + \sum_{j=1}^{m}\left(A'_j \operatorname{Cos}\frac{2\pi j\tau}{\omega} + B'_j \operatorname{Sin}\frac{2\pi j\tau}{\omega}\right) \quad ...(7)$$

Where, σ_τ is the evaluated standard deviation of monthly value; S_y is the standard deviation of series Y_t with t = 1, 2, 3,….., N (number of discrete values of rainfall); A'_j and B'_j are Fourier coefficients of standard deviation series; τ = 1, 2, 3,……, ω; j = 1, 2, 3,……, m. The parameters A'_j and B'_j were estimated by the equations:

$$A'_j = \frac{2}{\omega}\sum_{\tau=1}^{\omega} S_\tau \operatorname{Cos}\left(\frac{2\pi j\tau}{\omega}\right) \quad ...(8)$$

$$B'_j = \frac{2}{\omega}\sum_{\tau=1}^{\omega} S_\tau \operatorname{Sin}\left(\frac{2\pi j\tau}{\omega}\right) \quad ...(9)$$

$$S_\tau = \left[\frac{1}{n}\sum_{\omega=1}^{n}\left(Y_{p,\tau} - m_\tau\right)^2\right]^{1/2} \quad ...(10)$$

Where, S_τ is the standard deviation of individual rainfall values.

The number of significant harmonics was required to compute periodic component in the mean and standard deviation. Since, a limited number of harmonics is sufficient to explain the major part of variance of a periodic parameter, it is not necessary to estimate all harmonics. In general, it is found that first six harmonics of a periodic parameter for time series of any interval less than or equal to 30 days are significant and should be tested for significance (Mutreja, 1986). The maximum number of harmonics of periodic parameter to be fitted in the trend-free monthly rainfall series was taken as six for testing of the significance. The explained variance by j^{th} harmonic of periodic component is expressed as the ratio of $(A_j^2+B_j^2)/2$ to the variance of standard deviation of monthly data.

The periodic mean and standard deviation were calculated after determination of significant number of harmonics. The periodicity of the time series was removed from the trend-free time series by standardization with the value of μ_τ and σ_τ by using the equation:

$$Y'_{p,\tau} = \frac{Y_{p,\tau} - \mu_\tau}{\sigma_\tau} \quad ...(11)$$

Since the variable $Y'_{\tau,p}$ in eq. (11) is only an approximately standardized variable, its mean $\bar{Y}'$ and standard deviation S'_y were determined to compute the standardized stochastic component by the equation (Mutreja, 1986):

$$E_t = \frac{Y'_{p,\tau} - \overline{Y'}}{S'_Y} \quad ...(12)$$

Where, E_t is the stochastic component of the series.

This standardization makes the expectation of the mean and variance at any time t equal to zero and one, respectively; thus approximating a stationary condition. The series E_t was subjected to checks for stationary behaviour and model identification.

Modeling of Stochastic Component

The stochastic component of the time series E_t may have only dependent or independent stochastic component. The dependent component of stochastic series was modeled through AR family of models as it is stationary in nature. In AR process, the current value of the process is expressed as a weighed sum of past values plus the current shock. An AR process of order p, AR (p) is represented as:

$$S_t = \phi_{p,1} S_{t-1} + \ldots\ldots\ldots + \phi_{p,p} S_{t-p} + a_t \qquad \ldots(13)$$

Where, S_t is the stationary stochastic dependent component; ϕ_p is AR model parameter p (= 1, 2,……..p); a_t is the independent random effect at time t, or residuals. The recommended steps for fitting of model represented by Eq. (13) are identification, parameter estimation, and verification of model type, order and parameters.

Model Identification

Autocorrelation and partial-autocorrelation functions were used for identification of the AR model. Durbin (1960) proposed the following equation to calculate the partial autocorrelation function of lag K:

$$PC_{K,K} = \frac{r_k - \sum_{j=1}^{k-1} PC_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} PC_{k-1,j}\, r_j} \qquad \ldots(14)$$

Where, $PC_{k,k}$ and r_k are partial autocorrelation function and autocorrelation function at lag K, respectively.

$$PC_{k,j} = PC_{k-1,j} - PC_{k,k} PC_{k-1,k-j} \qquad \ldots(15)$$

Identification of AR models was done with the standard error of partial autocorrelation and autocorrelation functions using the equations proposed by Anderson (1976). The autocorrelation and partial autocorrelation functions for the series E_t were plotted against lag K and by visual inspection of the plot, tentative order of the model was selected.

Parameter Estimation

AR parameters were expressed in terms of autocorrelation functions and were computed by the general recursive formula (Kottegoda, 1980):

$$\phi_{p,p} = \left(r_p - \sum_{j=1}^{p-1} \phi_p - 1, r_{p-j} \right) \Bigg/ \left(1 - \sum_{j=1}^{p-1} \phi_{p-1,j}\, r_j \right) \qquad \ldots(16)$$

$$\phi_{p,j} = \phi_{p-1,j} - \phi_p, p\phi_{p-1,p-j} \qquad j = 1, 2, 3, \ldots, p-1 \qquad \ldots(17)$$

Diagnostic Checking

The Box-Pierce portmanteau lack of fit test was used to check the adequacy of the model. This test was applied collectively to a set of serial correlation r_1 of an estimated sequence of residuals. The order of AR model was determined by the method given by Jenkins and Watt (1968).

Independent Stochastic Component

After identification of the dependent stochastic component, it was separated from the series and the new series is called independent stochastic component with an independently distributed error term with zero mean and unit variance. Before modeling the independent stochastic component, the normality of the series was checked by the coefficient of skewness, since a normal distribution has a coefficient of skewness equal to zero.

A fully symmetrical distribution would exhibit the property that the coefficient of skewness is equal to zero. There are two alternative ways to provide a fit to skewed hydrologic data, one is to fit the available distribution functions to obtain a good fit and other is to transform the data in order to normalize them. It is often convenient to seek normalization by transformation. Box-Cox transformation was used to transform the skewed data to normal data. Box and Cox (1964) suggested the following method of transformation:

$$y = \frac{X^{\lambda} - 1}{\lambda} \; ; \; \lambda \neq 0 \qquad \ldots(18)$$

$$y = \ln x \; ; \; \lambda = 0 \qquad \ldots(19)$$

Where, y is the transformed value of variate x in the given series η_t. The parameter λ was chosen such that the skewness of the data series was the minimum. Hinkley (1997) suggested a simpler procedure for transformation. According to this procedure d_λ was calculated for each of the different value of λ varying from -1 to +1 with an interval of 0.1. The test statistic is express as:

$$d_{\lambda} = \frac{\overline{y} - \varsigma_y}{\sigma_y} \qquad \ldots(20)$$

Where, $\bar{y}$ is mean of the transformed series; $y\varsigma$ is median of the transformed series; and σ_y is standard deviation of the transformed series. The value of d_λ is to be compared and interpolated, if necessary, to find the value of λ for which d_λ is minimum. Generally, value of λ varies from -1 to +1 for rainfall series. The value of λ was assigned at an interval of 0.1 varying from -1 to +1 and statistic d_λ and the coefficient of skewness were calculated. The value of λ for which d_λ was minimum was used for the transformation of series into a normal series. The normal series was then expressed as:

$$Rt = \mu t + \sigma t\, \varepsilon t \quad ...(21)$$

Where, R_t is the transformed residual series and expressed as:

$$R_t = \frac{(\eta_t)^\lambda - 1}{\lambda} \quad ...(22)$$

Where, μ_t is mean of the residual series η_t; σ_t is standard deviation of series η_t; and ε_t is a random component with zero mean and unit variance (obtainable from standard tables). Finally, the mathematical model was developed as a sum of deterministic and stochastic component. The developed model was used to generate future sequences of rainfall data. To apply the model, the random numbers are required to be generated by using the table presented by Krumbein and Graybill (1965).

3. RESULTS AND DISCUSSION

The mathematical model for monthly rainfall time series was developed and evaluated for regeneration and prediction performances using statistical parameters measures such as mean absolute error (MAE), mean square error (MSE), mean forecast error (MFE), ISE, and WPM. Monthly rainfall data for the years 1975 to 2008 of the Rahuri rainguage station were used to develop the mathematical model. The observed monthly rainfall data were plotted against time as shown in Fig. 1. The historical data series of monthly rainfall was found to be non-stationary. Therefore, to smoothen the recorded monthly rainfall series, it was transformed by square root transformation using Eq. 1. The transformed monthly rainfall data series was divided into deterministic components and stochastic components and mathematical model was developed after identification and evaluation and removal of these components using the step-wise procedure.

Deterministic Components

The transformed monthly rainfall values indicate no trend present in the series X_t, as confirmed by the turning-point test and Mann-Kendall's rank-correlation test at 5% level of significance. The variance of total number of turning points in standard normal form (= -0.553) for turning point test, and that of Mann-Kendall's test (= 0.92) were well within the table value (±1.96) at 5% level of significance. Since, there was no trend component in the series X_t, the trend-free series X_t was expressed as series Y_t for further analysis.

The periodic component in the series was investigated by the method of harmonic analysis after determination of base period through correlogram analysis and significant number of harmonics. The autocorrelation functions r_k of the series Y_t for lag 1 to 180 were computed using Eq. 2. The autocorrelogram of series Y_t was developed by plotting autocorrelation

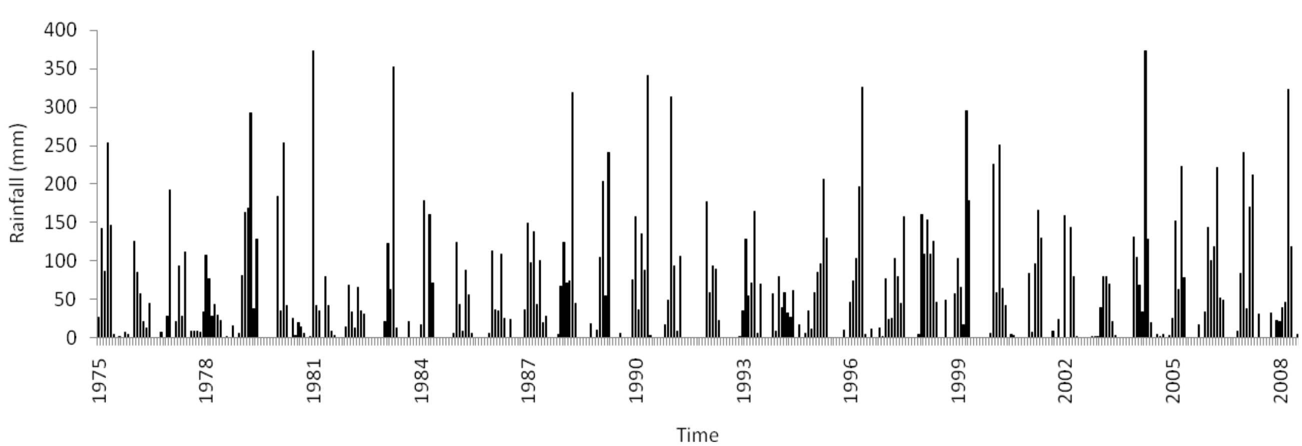

Fig. 1. Observed monthly rainfall for calibration period at Rahuri in Maharashtra, India

function with lags up to 48 as shown in Fig. 2, which revealed that autocorrelation functions are significantly different from zero, indicating that the monthly rainfall values are mutually dependent. It was also observed from the autocorrelogram that the narrow peak and broad troughs combining together make one cycle of a period of 12 months. Therefore, the series Y_t was characterized as having a periodicity of 12 months. The periodic mean μ_τ was determined by using Eq. 3 supported by Eqs. 4 through 6; and periodic standard deviation σ_τ was determined by using Eq. 7 supported by Eqs. 8 through 10, respectively. The values of coefficients A_j, B_j and A'_j and B'_j were computed upto six harmonics along with their explained variances. The actual number of harmonics to be fitted in the series Y_t were determined through an ANOVA test and it was observed that four harmonics are considered significant for determining the periodic component for further analysis.

After determination of significant harmonics, month wise periodic mean and standard deviations for twelve months were computed by Eqs. 3 and 7, respectively. The periodicity of the series Y_t was then removed by Eq. 11, and the components thus obtained were standardized with the help of Eq. 12 as suggested by Srikanthan and McMohan (1983). The series produced after standardization was termed as a dependent stochastic series S_t and its mean and standard deviation were computed month wise to check stationarity. Since, the computed values of mean and standard deviation do not change with time, *i.e.* the series properties are independent of the time. Therefore, the series S_t was considered as stationary.

Stochastic Components

Dependent structure of the stochastic component from trend- and periodicity-free series S_t was modeled through the auto regressive family of models following the steps such as identification of the model type and order, estimation of parameters, and diagnostic checking for the adequacy of selected model. Autocorrelation functions and partial autocorrelation functions are required for construction of autocorrelogram and partial autocorrelogram for identification of proper type and order of the AR model.

As suggested by Anderson (1976), the AR models up to order six were tried, as no improvement in residual sum of squares after order 6 was observed. Autocorrelation function of residuals of AR (1) was not significant. It was evident from the correlograms that the residuals obtained from all the five orders are mutually independent as they are not significantly different from zero. The test statistics was then compared with the table value of chi-squared and it was observed that [the test static (3.93) for model AR (6) was found to be less than the table value (5.99). Therefore, the model AR (6) gave a good fit and was acceptable. In order to fit the model AR (6), SAS-9.2 software was used. The parameters for the AR (6) model were determined by Eqs. 16 and 17. The estimated values of parameters were found to be $\phi_{6,1}$ = 0.1764, $\phi_{6,2}$ = 0.043, $\phi_{6,3}$ = - 0.032, $\phi_{6,4}$ = 0.085, $\phi_{6,5}$ = -0.071 and $\phi_{6,6}$ = 0.066. Therefore, stochastic component of the series was expressed as:

$$S_t = 0.1764\, S_{t-1} + 0.043\, S_{t-2} - 0.032\, S_{t-3} + 0.085\, S_{t-4} - 0.071\, S_{t-5} + 0.066\, S_{t-6} \quad ...(23)$$

The dependent stochastic component represented by AR family of model was separated from the series S_t. The independent stochastic component a_t was determined by AR (6) model using Eq. 23. The autocorrelation functions of the independent stochastic component were computed by Eq. 2 and the autocorrelogram up to lag 48 was plotted. The value of coefficient of skewness was found to be 0.379. Since the value of coefficient of skewness is not equal to zero it could be concluded that a_t series is not a normally distributed series. Therefore to obtain the normalized form, the Box-Cox transformation was applied. For

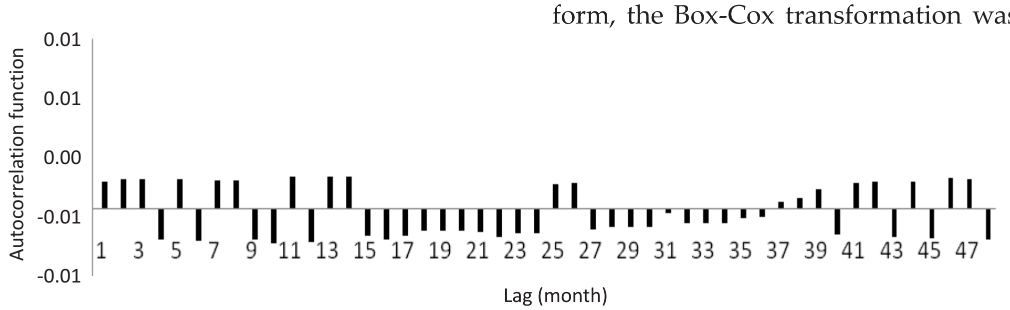

Fig. 2. Autocorrelogram for trend-free monthly rainfall time series Y_t

each value of λ the sample mean, standard deviation and median were computed and the statistic d_λ was determined by Eq. 20, which varied from -1 to 1 at an increment of 0.1. The values of d_λ and coefficient of skewness were found to be the minimum as 0.021 and 0.007, respectively at λ = 0.7. Therefore, a suitable value of Box-Cox transformation was found to be 0.7. After substituting transformed residual series in Eqs. 21 and 22, the R_t series was expressed as:

$$\frac{(\eta_t)^{0.7}-1}{0.7} = \mu_t + \sigma_t.\varepsilon_t \qquad \text{Or,}$$

$$\eta_t = [0.7(\mu_t + \sigma_t.\varepsilon_t)+1]^{1/0.7} \qquad \text{...(24)}$$

Eq. (24) represents the independent stochastic component. The mean μ_t and the standard deviation σ_t of the series R_t were found to be 1.109 and 0.946, respectively.

Finally, the mathematical model of the time series was obtained by substituting the values of periodic and stochastic components in Eq. 1 as:

$$Z_t = \begin{bmatrix} 4.77 + \begin{Bmatrix} (-2.67)Cos(\pi/6)\tau - 4.277Sin(\pi/6)\tau - 0.217Cos(\pi/3)\tau + 0.338Sin(\pi/3)\tau \\ -1.34Cos(\pi/2)\tau + 0.48Sin(\pi/2)\tau + 1.154Cos(2\pi/3)\tau + 0.33Sin(2\pi/3)\tau \end{Bmatrix} \\ + \{0.176S_{t-1} + 0.043S_{t-2} - 0.032S_{t-3} + 0.085S_{t-4)} - 0.071S_{t-5} + 0.066S_{t-6}\} \\ + [\{0.7(1.109 + 0.946\varepsilon_t) + 1\}^{1/0.7}\varepsilon_t] \end{bmatrix}^2 \qquad \text{...(25)}$$

This mathematical model of the time series was validated for regeneration for 1975 to 2008 and prediction of monthly rainfall data for 2009 and 2010. Mean monthly rainfall values of the historical and regenerated series for calibration period (1975-2008) are given in Table 1.

Table: 1
Statistical indices for observed and regenerated mean monthly rainfall series (1975 to 2008)

Month	Mean of observed series (mm)	Mean of regenerated series (mm)	MAE[1] (mm)	ISE[2] (%)	WPM[3]	CE[4]
January	2.3	5.8	3.5	3.53	0.76	0.97
February	3.1	4.6	1.5			
March	3.7	8.6	4.9			
April	5.5	10.0	4.5			
May	21.1	25.7	4.6			
June	110.8	117.1	6.3			
July	76.3	84.6	8.3			
August	82.6	89.2	6.6			
September	151.9	151.4	0.5			
October	78.0	90.9	12.9			
November	20.8	40.0	19.2			
December	8.9	14.9	6.0			
Mean	47.1	53.6	6.6			

[1]Mean absolute error; [2]Integral square error; [3]Weighted Pearson moment; [4]Coefficient of efficiency

Testing and Evaluation of the Model

To evaluate the performance, the model was tested for the calibration period (1975 to 2008) and it was also run for prediction period (2009 to 2010) of monthly rainfall data. Graphical comparison of autocorrelation functions of observed and regenerated series was made to assess the regeneration performance of the model. Autocorrelation functions of the observed and regenerated data series were computed by using Eq. 2. To evaluate the results of model graphically, the autocorrelation error plot indicated that the pertinent characteristics of time series are maintained and the properties propagated in the regenerated series are similar to that of the historical series.

Several types of statistical parameters were also used to measure the degree of agreement between the observed and regenerated as well as predicted values of monthly rainfall. The mean and standard deviation of the regenerated monthly series were found to be 53.57 mm and 50.71 mm, respectively, which were close to the values of 47.08 mm and 50.89 mm of the observed monthly series. The statistical error functions were also used for evaluation purpose as follows:

The MAE was computed for the rainfall time series as the mea of the absolute magnitude of difference between the regenerated and observed monthly rainfall values. The absolute errors for January to December of the regenerated series are given in Table 1. The value of absolute error for various months ranged from 0.5 to 19.2 mm.

Singh *et al.* (1991) used the ISE as a measure of goodness of a fit of time series model for air temperature. The ISE for the mathematical model was estimated by using the Eq. 26:

$$ISE = \frac{\sqrt{\sum_{i=1}^{m}[R_o - R_s]^2}}{\sum_{i=1}^{m} R_o} \times 100 \qquad \text{...(26)}$$

The value of ISE for the regenerated series was found to be 3.53%, which is fairly within the acceptable limits (Table 1).

Weighting the Pearson product-moment correlation coefficient by the ratio of standard deviation was used as a simple index for comparing the specific events (McCuen and Snyder, 1975). The WPM was computed by the Eq. 27:

$$WPM = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(R_s - \overline{R}_s\right)}{S_{Rs}}\frac{\left(R_o - \overline{R}_o\right)}{S_{Ro}} \quad ...(27)$$

Where, $\hat{R}_s$ is the mean of regenerated monthly rainfall; $\hat{R}_o$ is the mean of observed monthly rainfall; S_{Rs} is the standard deviation of simulated monthly rainfall; S_{Ro} is the standard deviation of observed monthly rainfall; and N is the number of observations. The weighted moment C is expressed as $C = S_{Ro}/S_{Rs}$, if $S_{Ro} < S_{Rs}$; or $C = S_{Rs}/S_{Ro}$, if $S_{Rs} > S_{Ro}$. The WPM is always smaller than the Pearson moment and approaches unity with perfect agreement between R_s and R_o. The WPM for the regenerated series was found to be 0.76 (Table 1) indicating a good fit.

The CE was also computed using the Eq. 28 (Nash and Sutcliffe, 1970):

$$CE = 1 - \frac{\sum_{1}^{N}(R_o - R_s)^2}{\sum_{1}^{N}(R_o - \overline{R_o})^2} \quad ...(28)$$

The CE between the observed and regenerated series was found to be 0.97 which indicates a good fit (Table 1). The statistical indices and for mean monthly rainfall values of observed and regenerated series (using Eqs. 26 through 28) are in given in Table 1, which indicates that the regenerated values, though under-predicted for all months, are reasonably close to the observed values.

Evaluation of Prediction Performance

The developed mathematical model for monthly rainfall time series was used for the prediction of monthly rainfall values of two years ahead *i.e.* for the years 2009 and 2010. The mean and standard deviation of the predicted monthly series were found to be 38.68 mm and 49.93 mm, respectively, which were close to the values of 39.40 mm and 52.91 mm of the observed monthly series. The forecast was made for all calendar months by taking the value of random component ε_t equal to zero (or from standard tables). The predicted values of monthly rainfall for the years of 2009 and 2010 are shown in Table 2, which indicates that the predicted values are reasonably close to the observed values. The monthly rainfall was slightly under predicted for June month onwards, while it was over-predicted for previous months of the year.

In order to evaluate the prediction performance of the model, the absolute error alongwith the ISE, WPM and CE were computed using Eqs. 26 through 28, respectively as given in Table 2. The error values are well within the acceptable limits. The CE between the observed and predicted series was found to be 0.99 which indicates a very good fit for prediction purpose.

Table: 2
Statistical indices for observed and predicted mean monthly rainfall series (2009 and 2010)

Month	Mean of observed series (mm)	Mean of predicted series (mm)	MAE[1] (mm)	ISE[2] (%)	WPM[3]	CE[4]
January	0	0.8	0.8	1.80	0.94	0.99
February	0	1.4	1.4			
March	0	5.7	5.7			
April	0.8	6.3	5.5			
May	3.4	6.2	2.8			
June	74.9	65.8	9.1			
July	161.3	156.7	4.6			
August	82.8	79.7	3.1			
September	92.9	89.3	3.6			
October	47.3	44.9	2.4			
November	7.1	5.9	1.2			
December	2.3	1.5	0.8			
Mean	39.4	38.7	3.4			

[1]Mean absolute error; [2]Integral square error; [3]Weighted Pearson moment; [4]Coefficient of efficiency

4. CONCLUSIONS

The following conclusions were drawn from the results of this study:

- The deterministic trend component for the observed monthly rainfall series was non-significant at 95% confidence level. However, the Fourier series analysis indicated that the periodic component consists of four significant harmonics.
- The MAE, ISE, WPM and CE for the regenerated and prediction series were well within the acceptable limits. It suggests the adaptability of the model for forecasting of monthly rainfalls at Rahuri region in Maharashtra, India.

REFERENCES

Anderson, O.D. 1976. *Time Series Analysis and Forecasting, the Box Jenkins Approach*. Butterwarth and Co., London.

Box, G.E.P. and Cox, D.R. 1964. An analysis of transformations. *J. Royal Stat. Soc.*, 26(2): 211-252.

Dabral, P.P., Pandey, A. Baithuri, N. and Mal, B.C. 2008. Stochastic Modelling of Rainfall in Humid Region of North East India. *Water Resour. Manage.*, 22: 1395-1407.

Durbin, J. 1960. *The fitting of time series models*. Review of the International Press Ltd., London, pp 233-244.

Gupta, M. and Singh, R. 2004. Stochastic modeling of evaporation in arid region of India. *Hydrol. J. (IAH)*, 27(3-4):1-13.

Haan, C.T., Allen, D.M. and Street, J.O. 1976. A Markov chain model of daily rainfall. *Water Resour. Res.*, 12(3): 443-449.

Haltiner, J.P. and Salas, J.D. 1988. Development and testing of multi variative seasonal ARMA (1,1) model. *J. Hydrol.,* 104: 247-272.

Hinkley, D. 1997. On quick choice of power transformation. *Appl. Stat.*, 26: 67-69.

Jayawardena, A.W., Xu, P.C., Tsang, F.L. and Li, W.K. 2006. Determining the structure of a radial basis function network for prediction of nonlinear hydrological time series. *Hydrol. Sci. J.,* 51(1): 21-44.

Jenkins, R.H. and Watt, D.G. 1968. *Spectral Analysis and Its Applications*, Holden Day, San Franscisco, California.

Khatua, S.K., Panigrahi, B. and Panigrahi, K. 2014. Probability analysis of maximum daily rainfall for hydrological design of soil and water conservation structures. *J. Indian Water Resour. Soc.,* 34(4): 9-16.

Kottegoda, N.T. 1980. *Stochastic Water Resources Technology*. The McMillan Press Ltd., London.

Kottegoda, N.T. and Horder, M.A. 1980. Daily flow model based on rainfall occurrences using pulses and transfer function. *J. Hydrol.*, 47: 215-234.

Krastanovic, P.F. and Singh, V.P. 1991. A univariate model for long-term stream flow forecasting. *Stoch. Hydrol. Hydrau.*, 5: 173-188.

Krumbein, W.C. and Graybill, F.A. 1968. *An introduction to statistical models in Geology*. McGraw Hill Book Co., New York.

Kumar, A., Pal, R. and Sharma, H.C. 2007. Probability analysis of monsoon rainfall data of Saharanpur for Agricultrual planning. *Indian J. Soil Cons.,* 36(3):131-135.

Louck, D.P., Stedinger, J.R. and Haith, D.A. 1981. *Water Resources Systems Planning and Analysis*. Prentice-Hall, Englewood Cliffs.

McCuen, R.H. and Snyder, W.M. 1975. A proposed index for comparing hydrographs. *Water Resour. Res.,* 11(6): 1021-1024.

Mutreja, K.N. 1986. *Applied Hydrology.* Tata McGraw Hill Publishing Company Limited, New Delhi, 959 p.

Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, part-I-A discussion on principles. *J. Hydrol*., 10 (3): 282-290.

Panigrahi, B. 1998. Probability analysis of short duration rainfall for crop planning in coastal Orissa. *Indian J. Soil Cons.,* 26(2): 178-182.

Panigrahi, B. and Panigrahi, K. 2013. *Engineering Hydrology*. New India Publishing Agency, New Delhi, 667 p.

Patra, K.C. 2008. *Hydrology and Water Resources Engineering*. Narosa Publishing House, New Delhi.

Raghuvanshi, N.S. and Wallender, W.W. 2000. Forecasting daily evapotranspiration for a grass reference crop. *J. Agril. Eng*., 9(1): 1-16.

Richardson, C.W. 1978. Generation of precipitation over an area. *Water Resour. Bull.*, 14(5): 1035-1047.

Richardson, C.W. 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resour. Res.*, 17(1): 182-190.

Sapolia, S.K., Chander, S. and O'Connor, K.M. 1980. An autocorrelation approach for parameter estimation of fractional order equal root autoregressive models using hyper geometric functioning. *J. Hydrol.*, 47: 1017.

Sen, Z. 1980. Effect of periodic parameters on autocorrelation structure of hydrologic series. *Water Resour. Res*., **15**(60): 1639-1642.

Singh, A.K., Singh, J. and Gupta, A.P. 1991. Time series analysis of air temperature of Udaipur. *Hydrol. J. IAH*, 14(5): 16-23.

Srikanthan, R. and McMohan, T.A. 1983. Stochastic simulation of daily rainfall for Australian Stations. *Trans. ASAE*, 26(3): 754-759.

Tularam, G.A. and Ilahee, M. 2010. Time series analysis of rainfall and temperature interactions in coastal catchments. *J. Mathe. Stat.,* 6(3): 372-380.