International Journal of Agricultural Engineering, Vol. 3 No. 1 (April, 2010) : 175-178

Research Paper :

Stochastic approach for monthly surface runoff variables of hilly watershed

MOHAMMAD GUFRAN

Accepted : *April, 2010*

Correspondence to:
MOHAMMAD GUFRAN
Department of Soil and Water Conservation Engineering, College of Agricultural Engineering and Technology, ETAWAH (U.P.) INDIA

ABSTRACT

A time series can be either stationary or non-stationary. The stationary property of a time series is diagnosed by evaluating its statistical properties. The statistical properties of the different components of a stationary time series do not change, except owing to sampling variations. In annual series of runoff data, when not affected by variations in climatic or watershed properties, is called a stationary time series data. In a non-stationary time series, the statistical properties change from one part to another, the data are time dependent, e.g. the runoff pattern in a year changes from one season to another. The stationary characteristics of a time series, based on its mean and variance properties, are accepted for modeling. In case of a stationary series, the mean, which is the first movement, is constant. The second movement is the covariance, which when divided by the variance gives the correlation. The theoretical correlation, known as autocorrelation, expresses the dependence of the time series data on each other. In present study monthly runoff data of mountainous watershed was collected and computed Fourier Coefficient of mean and standard deviation and also removal of periodicity. Periodic service of mean was found A and B, -33.92 and -1251.12, respectively. Similarly A_2, B_2, A_3 and B_3 were found, -242.02, +100.64, 267.49 and -355.01, respectively. Study also revealed that periodic service of standard deviation was estimated as A_1, B_1, A_2, B_2, and A_3, B_3 were 1449.56, -1862.22, 582.51, -1540.21 and 97.79, -1716.32, respectively.

Key words : Autocorrelation, Variance, Stationary and non stationary, Mountainous, Runoff, Forecasting

A time series is a set of observed data recorded at specified times, usually spaced at equal intervals. Mathematically, a times series can be expressed by the values Z_1, Z_2, Z_3, etc of a variables Z at times t_1, t_2, t_3, etc.

A time series model can be divided into two components *viz*., deterministic and stochastic components. The deterministic component is used for prediction of the time and chance independent future events, while the stochastic component is used for the determination of the chance and chance dependant effects.

Deterministic components are either periodic or non-periodic in nature. The non-periodic component is characterized by its trend and jump characteristics. Trend characteristics could be either of an increasing type or of a decreasing type. A variation in trend characteristics of the data series is caused mainly because of change in the physical characteristics of the watershed and in the case of jump, it is due to changes caused by a sudden variation in any of the watershed characteristics. The periodic nature of a deterministic component is characterized by its cyclic pattern which exhibits an oscillatory movement and is repeated over a fixed interval of time.

Stochastic models always have outputs that are variable in time. They may be classified as time independent or time correlated. A time independent model represents a sequence of hydrologic events that do not influence each other, while a time-correlated model represents a sequence in which the next event is partially influenced by the current one and possibly by other in the sequence.

The present study is based on analysis of stochastic components of monthly runoff data of mountainous watershed and determined the stationary and non-stationary series.

METHODOLOGY

The suitable models for the solution of stochastic components of the time series, its stationary behaviour is checked first. Variations in stationary behaviour of the S_t series may occur because of seasonal effects. To conduct checks for stationary behaviour, the monthly mean and the standard deviation of each month data were computed. If it was found that the mean values did not fluctuate around zero, and the standard deviation was not changing in all the months, then the time series was taken as stationary without seasonality effects, otherwise, it was considered to have seasonal effects. Suitable models for each case were then selected.

Stationary series:

The depended stochastic components (D_t) of the S_t series was modeled by the ARMA family of models, when the series was stationary in nature, an iterative procedure for the fitting of ARMA models was used. This was performed recursively. The essential steps comprised identification, parameter estimation, and verification of the model type, order and parameter.

ARMA models:

Autoregressive moving average (ARMA) models are linear stochastic models. They are analogous to conceptual models of parametric hydrology, based on linear reservoirs. ARMA models can not be estimated exactly as they were constituted by several random effects. Linear stochastic models are selected to forecast data of one or more time periods ahead, and to generate a synthetic data sequence of the time series.

Autoregressive process of order p – A R(p) model, is delineated as:

$$S_t = \phi_{p-1} S_{t-1} + \ldots\ldots\ldots\ldots + \phi_{p-p} S_{t-p} + R_t \quad \ldots\ldots(i)$$

$$= \sum_{k-1}^{p} \phi_{pk} S_{t-k} + R_t \quad \ldots\ldots.(ii)$$

In the AR(p) process, the current value of the process is ellucidated as a weighted sum of the past values, plus the current shock.

Autoregressive moving average ARMA (p,q) models:

A reasonable expansion to AP (p) and MA (q) models is a mixed model of the form:

$$S_t = \phi_{p-1} S_{t-1} + \ldots\ldots\ldots\phi_{p.p} S_{t-p} + R_t - \theta_{q.1} R_{t-1} - \theta_{q.q} R_{t-q} \quad \ldots\ldots.(iii)$$

$$= \sum_{k-1}^{p} \phi_{p.k} S_{t-k} + Rt - \sum_{k-1}^{q} \theta_k R_{t-k} \quad \ldots. (iv)$$

where :

S_t = Stationary stochastic component of the original series

φ_k= Autoregreassive model parameter, K= 1, 2 ..., p

θ_k= Moving average model parameter, K= 1,2, ..., q

p= Order of autoregressive process

q= Order of moving average process

R_t= Independent random effect at time *t* of residuals

The mixed model is called the ARMA (p,q) model.

Computation of stochastic components:

The dependent stochastic component D_t when separated from the stochastic S_t gives the independent stochastic components as:

$$S_t = D_t + R_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (v)$$

Therefore, $S_t = D_t + R_t$

In equation (iv), if the component R_t is deducted from both sides then by applying equation (v) the remainder is equal to D_t and is ellucidated as:

$$D_t = \sum_{k-1}^{p} \phi_{pk} S_{t-k} - \sum_{k-1}^{q} \theta R_{t-k} \quad \ldots\ldots (vi)$$

This gives the equation for the dependent stochastic component. The values of the independent stochastic component R_t is obtained by deducting the value of D_t from S_t. The R_t series is then subject for diagnostic checking, or verification, by testing it for independence.

Parameter estimation:

The following recursive formula is used to compute the AR parameters (Kottagoda, 1980):

$$\phi_{p-p} = \frac{r_p - \sum_{j=1}^{p-1} \phi_{p-1.j} r_{p.j}}{1 - \sum_{j=1}^{p-1} \phi_{p-1.j} r_j} \quad \ldots\ldots\ldots \quad (viii)$$

$$\phi_{p-1} = \phi_{p-1.j} - \phi_{p.p} \phi_{p-1} \phi_{p-j}$$

j=1,2,3.......................p-1

where, φ is the AR parameter and p is the order of the autoregressive process.

The moving average parameter j is computed by the following formula of Anderson (1976)

$$r_k = \frac{\theta_k + \sum_{j=1}^{q-k} (\theta_j)(\theta_{k+k})}{1 + \sum_{j=1}^{q} \theta_j^2} (k = 1,2\ldots\ldots\ldots, q) \quad (viii)$$

It is not necessary that these be the best estimates. Alternatively, the parameter is estimated by minimizing the sum of squares of the residuals.

Non-stationary series:

For a non-stationary series, seasonal models are selected. Yevjevich (1972) presented the following generalized seasonal model to compute the stochastic component which are given below:

$$S_t = S_{t.b} = \mu_b g_b \sum_{q=1}^{b} \phi_{p.q} \frac{(S_{t-q.b-q} - \mu_{b-q})}{g_{b-q}} + g_b A_e \quad (ix)$$

where

$S_{t.b}$= non-stationary data for period *t*, where *t*= 1,2,........... 1,2, R (R is the number of the year) .

m_b= mean of the month b

j_b = Standard deviation of the month b

A_e = non stationary independent random component.

p= order of the model

q = 1,2, 3,............................., b

$$A_e = \frac{(S_{t.b} - x_{t.b})}{g_b} \qquad \text{(x)}$$

From equation (x), the value of A_e is determined. To obtain the value of A_e parameters of seasonal models are computed by using equations for estimation of parameter for ARMA models.

The remaining series represented by A_e is then subjected to diagnostic checking or verification by testing it for independence, as done earlier for R_t of the stationary series.

The results obtained to the system response function determined based on the mathematical formulation, adequacy of the mathematical model developed in this study for computing runoff data to represent the mountainous watershed through autocorrelation function and related standard errors are analysed.

Synthetic data were generated and R_t series was checked for its frequency distribution. Generally, normal, log-normal or Pearson type distribution was fitted to the R_t series. Alternatively, the R_t series was transformed to normal by power transformation. The normality of the series is checked by chi-sequence test of Kolmogorov – Smirnow goodness of fit test or by plotting its cumulative distribution function on a normal probability paper. A normal series plots as a straight line and the transformed R_t series in terms of the A_t series, is expressed as:

$$A_t = «_t + ¢_t s_t \quad .... \qquad \text{(xi)}$$

where

A_t = Transformed R_t series

μ_t = Mean of the A_t series

δ_t = Standard deviation of A_t series

ς_t = Random component with zero mean and unit variance.

The developed sub-models for component of the decomposition model were added together to get the final model of the Z_t series. The Z_t model could be used either for data generation by taking the values of standard normal variants ς_t from the standard tables or for short term forecasting by assuming the standard normal ς_t =0

Thus the expression for forecasted value is ellucidated below :

$$Y_t(L) = T_{t+L} + P_{t+L} + S_{t+L} \quad .... \text{(xii)}$$

where, $Y_t(L)$ is the forecasted value of the time t+L. The T_{b+L}, P_{t+L} and S_{t+L} are the respective values of trend periodic and stochastic components at the time t+L.

RESULTS AND DISCUSSION

The results obtained to the system response function determined based on the mathematical formulation, adequacy of the model developed in this study for computing runoff data to represent the mountainous watershed through autocorrelation functions and related mean and Fourier coefficients of first, second and third harmonic A_1, B_1, A_2, B_2 and B_3, B_3 of periodic service of mean were summarized in Tables 1 and 2 is given below.

The total variance is 37.64% explained by the entire three harmonic. It may be conducted that the third harmonic is 9.95% of the variance. The computed values

Table 1 : Fourier coefficients of different harmonic of period series of mean deviation

Values of	A_1	B_2
Fourier coefficients of first harmonic of periodic series of mean	-203.54	7506.73
Values of	A_2	B_2
Fourier coefficients of second harmonic of periodic series of mean	-1452.17	603.85
Values of	A_3	B_3
Fourier coefficients of third harmonic of periodic series of mean	1604.94	-2130.10

Table 2 : Fourier coefficients of different harmonic of period series of standard deviation

Values of	A_1	B_2
Fourier coefficients of second harmonic of periodic series of standard deviation	8697.39	11117.32
Values of	A_2	B_2
Fourier coefficients of second harmonic of periodic series of standard deviation	3171.80	-9241.29
Values of	A_3	B_3
Fourier coefficients of second harmonic of periodic series of standard deviation	586.78	-10297.97

of Y=0.41 and Sy=2.1 for the monthly average discharges after the removal of periodicity.

Conclusion:

The main objectives of this study were to develop time series analysis approach for monthly runoff data of mountainous watershed. The conclusions drawn from the results of the investigations are discussed below

– The computed values of Fourier coefficients of first, second and third harmonic of periodic series of mean are A_1, B_1, A_2, B_2 and A_3, B_3, (-203.54, -7506.73, -1452.17, 603.85, and 1604.98, -2130.10).

– The computed values of Fourier coefficients of first, second and third harmonic of periodic series of standard deviation are A_1, B_1, A_2, B_2 and A_3, B_3, (8697.39, 11117.32, 3171.80, -9241.29 and 586.78, -10297.97).

– The percentage of explained variance by the third harmonic was 6.59%. The first two harmonics explained 54.56% while the first three harmonics explained 61.15% of the variance. Hence only three harmonics were considered for the periodic series of mean.

– The total variance explained by the entire three harmonics was 37.64%. It may be noted that the third harmonic was 9.95% of the variance.

The monthly average discharge after removal of periodicity was only appox. a standardized variable, its mean 'Y' and standard deviation 'S_y' have bean computed as Y= - 0.41 and Sy=2.01

REFERENCES

Anderson, R.L. (1976). Distribution of the serial correlation coefficient. *Annals Mathematical Statistics*, **13** : 1-13.

Yevjevich, V. (1972). Probability and statistics in hydrology (Water resources publications, Fort Collins, Colorado, U.S.A.)

——— *** ———