STOCHASTIC MODELLING OF RAINFALL AT UDAIPUR

S. R. Bhakar and Raj Vir Singh
Deptt. of SWE, CTAE, Udaipur – 313 001 Rajasthan
(E-mail: srbhakar@rediffmail.com)

ABSTRACT

Stochastic modelling for rainfall has done using 100 years (1901–2000) data. The performed statistical tests indicated that the series of the monthly rainfall data was trend free. The periodic component of monthly rainfall could be represented by third harmonic expression. The stochastic components of the monthly rainfall followed second order Markov model. Validation of generated monthly rainfall series was done by comparison of generated and measured series. The correlation coefficient between generated and measured rainfall series was found to be 0.9974. The correlation was tested by t-test and found to be highly significant at 1 per cent level. The regression equation is very near to 1:1 line. Therefore, developed model could be used for future prediction of monthly rainfall.

INTRODUCTION

Rainfall is an important weather parameter for estimation of crop water requirements. Frequently, it is required to estimate rainfall of places where measured rainfall data are not available. Rainstorms vary greatly in space and time. This type of series could be estimated by the sum of periodic series and stochastic series. Periodic component takes into account the portion, which repeat after certain duration. The stochastic component constituted by various random effects, can not be estimated exactly.

MATERIALS AND METHODS

Location of the Study Area

The study was conducted at the Department of Soil and Water Engineering, College of Technology and Engineering, MPUAT, Udaipur. The area comes under the sub-humid region of the agro-climatic zone IV-A of the state of Rajasthan, and is situated at 24°35' N latitude, 73°42'E longitude and at an altitude of 582.17 m above mean sea level.

The climate of an area is characterized by variation in the climatological parameters over the years and from one year to the next. The annual rainfall in this region is 646.6 mm and more than 80% of this amount is received during the monsoon season alone, due to the influence of the southwest monsoon.

Collection of Rainfall Data

The rainfall data of Udaipur were collected from Meteorological Observatory of the College of Technology and Engineering, Udaipur. Meteorological data for a period of 100 years (1901-2000) were used in the study.

The mathematical procedure adopted for formulation of a predictive has been discussed in the following sub-sections: The principal aim of the analysis is to obtain a reasonable model for estimating the generation process and its parameters by decomposing the original data series into its various components. Generally a time series can be decomposed into a deterministic component, which could be formulated in manner that allowed exact prediction of its value, and a stochastic component, which is always present in the data and can not strictly be accounted for as it is made by random effects. The time series, $X_{(t)}$, was represented by a decomposition model of the additive type, as follows:

$$X_{(t)} = T_{(t)} + P_{(t)} + S_{(t)} \tag{1}$$

where,

$T_{(t)}$ = trend component, t = 1,2 ....N
$P_{(t)}$ = periodic component
$S_{(t)}$ = stochastic component, including dependent and independent parts.

To obtain the representative stochastic model of time series, identification and detection of each component of Equation (1) was necessary. A systematic identification and reduction of each component of $X_{(t)}$ was done, procedures of which are described below:

Trend Component

The trend component describes the long smooth movement of the variable lasting over the span of observations, ignoring the short term fluctuations. The basic idea here was to study only $T_{(t)}$ while eliminating the effects of other components. It leads to use the total seasonal data, Z_t, for identification of $T_{(t)}$ so that other components were suppressed. For detecting the trend, a hypothesis of no trend was made. Turning point test and Kendall's rank correlation tests as suggested by Kottegoda (1980), were performed: If the calculated value of z is within its table value, then, it can be concluded that the trend is not present in the series.

Periodic Component

The periodic component concerns an oscillating movement which is repetitive over a fixed interval of time (Kottegoda, 1980). The existence of $P_{(t)}$ was identified by the correlogram, a plot of autocorrelation coefficients, r_l, versus lag l. The oscillating shape of the correlogram verifies the presence of $P_{(t)}$, with the seasonal period P, at the multiples of which peak of estimation can be made by a Fourier Analysis followed by the tests for significant harmonics. The correlogram of the time series clearly show the presence of the periodic variations indicating its detection. The periodic component ries $X_{(t)}$ was expressed in Fourier form as follows:

$$X_{(t)} = A_o + \sum_{K=1}^{\infty}\left[A_K \operatorname{Cos}\left(\frac{2\pi Kt}{p}\right) + B_K \operatorname{Sin}\left(\frac{2\pi Kt}{p}\right)\right] \tag{2}$$

where,

$$A_o = \frac{1}{N}\sum_{t=1}^{N} x_{(t)} \tag{3}$$

$$A_K = \frac{2}{N}\sum_{t=1}^{N} x_{(t)} \, \text{Cos}\left(\frac{2\pi Kt}{p}\right) \tag{4}$$

and

$$B_K = \frac{2}{N}\sum_{t=1}^{N} x_{(t)} \, \text{Sin}\left(\frac{2\pi Kt}{p}\right) \tag{5}$$

where,

K = number of significant harmonics, p = base period, N = number of observation points and A_K and B_K = Fourier coefficients.
These coefficients were obtained by a least square fit of the data to the K^{th} harmonic components, then a least squares approximation could be given by the finite series.

In Equation (6) if $M \rightarrow \infty$, $P_{(t)} \rightarrow X_{(t)}$, then $X_{(t)}$ could be represented satisfactorily by Equation (6) only. However it may not be practical or desirable to allow the condition, $M \rightarrow \infty$. Thus the appropriate approach would be the selection of a value of M which contains only those harmonics which are significantly contributing towards $X_{(t)}$. With this as the objective two tests namely Test of analysis of variance and Fourier decomposition of mean square were conducted.

Stochastic Component

The stochastic component was constituted by various random effects, which could not be estimated exactly. A stochastic model in the form of Autoregressive model, AR, was used for the presentation of the time series. In this model, the current value of the process was expressed as a finite, linear aggregate of values of the process and a variate that was completely random. This model was applied to the $S_{(t)}$ which was treated as a random variable i.e. deterministic components were removed and the residual was stationary in nature. Mathematically, an autoregressive model of order p, AR (p) can be written as:

$$S_{(t)} = \sum_{K=1}^{P} \phi_{P,K} \, S_{(t-K)} + a_{(t)} \tag{6}$$

$$= \phi_{p,1} \, S_{(t-1)} + \phi_{p,2} S_{(t-2)} + \ldots.. + \phi_{p,p} S_{(t-P)} + a_{(t)} \tag{7}$$

where,

$\phi_{P,K}$ = autoregressive model parameters, K = 1,2 ...,p.
$a_{(t)}$ = independent random number

The fitting procedures of the AR(p) model to the meteorological parameters series involved selection of order (p) of the model.

Estimation of autoregressive parameters: The parameter estimation deals with the estimation of the autoregressive parameters of Equation (7). These parameters could be expressed in terms of serial correlation coefficient, as Yule-Walker equations (Bhakar, 2000). The general recursive formulae for estimating these parameters ($\phi_{p,k}$), where suffix p and k indicate the order and the number of parameter of the order in AR (p) model, respectively could be written as follows:

$$\phi_{p,k} = \left[\frac{r_p - \sum_{k=1}^{p-1} (\phi_{p-1,k})(r_{p-k})}{1 - \sum_{k=1}^{p-1} (\phi_{p-1,k})(r_k)} \right] \quad (8)$$

and

$$\phi_{p,k} = \phi_{p-1,k} - \phi_{p,p} \cdot \phi_{p-1,p-k}\, ; k = 1,2,3....p-1 \quad (9)$$

In Equation (8) r_k is the autocorrelation coefficient, Auto Correlation Coefficient of the series for K and was computed, for any series $Y_{(t)}$ at any lag, *l*, as follows:

$$r_l = \sum_{t=1}^{N-l} [Y_{(t)} - \overline{Y}_{(t)}][Y_{(t+l)} - \overline{Y}_{(t)}] / \sum_{t=1}^{N} [Y_{(t)} - \overline{Y}_{(t)}]^2 = C_l / C_o \quad (10)$$

where,

$Y_{(t)}$ = mean of the series, $Y_{(t)}$
N = total number of discrete values of $X_{(t)}$
C_l = autocovariance function at lag l, l= 0, 1, ...p

After estimating the AR parameters $\phi_{p,k}$, $S_{(t)}$ was calculated by using Equation (7). The sum of the periodic and stochastic component forms the generated value of the observed data. The difference was termed as residuals which were tested to check the adequacy of the formulated model.

Diagnostic Checking of the Model

Diagnostic checking concerns the verification for the adequacy of the fitted model. The examination of the autocorrelation structure of the residuals provides a powerful way of diagnostic checking. The residuals were examined for any lack of randomness. If the residuals were not random or were autocorrelated, the model has to be modified until the residuals become uncorrelated.

RESULTS AND DISCUSSION

For testing the statistical characteristics of monthly rainfall series 100 years data of monthly rainfall was taken Average of 100 was taken for every month to get mean monthly rainfall series. The statistical characteristics of the mean monthly rainfall series were estimated. Mean monthly rainfall values vary from 4.0 mm in December to 203.8 mm in July. Mean value was found to be 53.16 mm. There is large variability among the monthly values of rainfall of different years. This was further confirmed by the estimated deviations. The standard deviation in the monthly rainfall values ranged from 7.629 to 115 mm during entire year. The variation may be attributed towards the natural changes in yearly climate. The variation in coefficient of variance ranges from 58 to 13141. This signifies the importance of variability of monthly rainfall series. Since the values of variance significantly different from zero, it confirms that rainfall is mutually dependent.

Serial Correlation Coefficient

The lag one serial correlation coefficient of observed series was calculated by using Equation (10) and was found to be 0.420. The respective confidence limits were estimated as 0.208 to -0.247 (Kottegod, 1980). The value of lag one serial correlation coefficient lies outside the range of confidence limits and is significantly different from zero. This again confirms that rainfall process is mutually dependent. From the analysis of coefficient of variation and serial correlation, it is confirmed that rainfall process is time variant and not an independent one. Thus the rainfall time series could be modeled on stochastic theory. The mutual dependence of the observed rainfall series was also confirmed by the correlogram (Fig 1).

Trend Component

For identification of trend components, annual rainfall series was used. The annual rainfall series was obtained by transforming the 100 years annual series. For detection of trend the hypothesis of no trend was made and value of test statistic (Z) was calculated by Turning Point Test and Kendall's Rank Correlation Test. The calculated values of test statistic (Z) are 0.15 and –14.30 for Turning point test and Kendall's Correlation Test respectively. The estimated value of test statistic (Z) obtained for Turning Point Test, and Kendall's Rank Correlation Test was within the 1 per cent level of significance. Hence the hypothesis of no trend was accepted. Further from the turning point test total numbers of turning points in series were found to be 72. This indicates that the rainfall series is random. From the above analysis it is confirmed that the trend component in rainfall time series is absent and the observed series may be treated as trend free series.

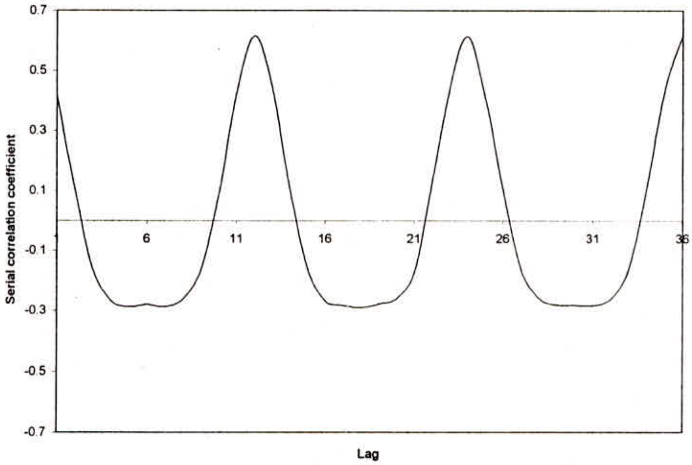

Figure 1 Correlogram of annual rainfall for Udaipur region

Periodic Component

To confirm the presence of periodic component in monthly rainfall series a correlogram was drawn. A correlogram is a graphical representation of serial correlation coefficients (r_t) as function of lag. l in which the value of r_l are plotted against respective value of l (Figure 1) The resulting oscillating shape of the correlogram confirms the presence of periodic component in the monthly rainfall. Further, the correlogram has peaks at legs equal to 12 and at other multiples of it. The time span of periodicity was taken as 12 for use in harmonic analysis of periodic component.

Determination of significant harmonics

For representing the periodic component of the rainfall series the numbers of significant harmonics were determined by analyzing by cumulative periodogram. Only first three harmonics are highly significant. Other harmonics were not significant and therefore could be ignored.

Parameters of periodic component

Using Equations (3), (4) and (5) the Fourier coefficients A_k and B_k were estimated. These Fourier decompositions for monthly rainfall series are also presented in Table 1. Study of Table 1 reveals that the first three harmonics explain more than 60 per cent of variance.

Table 1. Fourier decomposition of periodic component of rainfall series at Udaipur

Order	A_k	B_k	Amplitude	Explained variance	Cumulative explained variance	Cumulative periodogram
1	-48.751	-67.92	83.605	41.68	41.68	0.685
2	-15.263	49.08	51.398	15.753	57.433	0.948
3	19.351	-8.028	20.95	2.617	60.05	0.993
4	-4.062	0.229	4.068	0.099	60.149	0.995
5	3.679	-3.509	5.084	0.154	60.303	0.997
6	-5.468	-3.028	6.251	0.233	60.536	1.000

Cumulative periodogram test

In this test a graphical procedure has been employed as criteria for obtaining the significant harmonics to be fitted in a periodic component. A graph was drawn between P_i and number of harmonics, called the cumulative periodogram. The fast increase in Pi has been considered as a significant harmonics and the rest of harmonics were rejected.

Estimates of the mean square deviation and the cumulative periodogram Pi for monthly rainfall series were made and the plot of Pi against i is shown in Figure 2. From the cumulative periodogram in the monthly rainfall series, it can be observed that the first three harmonics appeared to be the periodic part of the fast increase and after that periodogram remains almost constant which may be treated as non-significant.

The two criteria used to identify the number of significant harmonics to be used in modelling periodic component were found to be consistent, so first three harmonics are treated as significant contributing towards periodicity and remaining are considered as white noise.

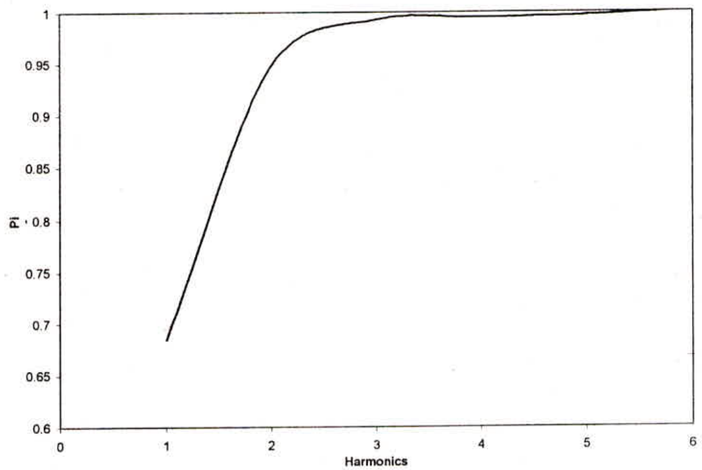

Figure 2 Cumulative periodogram of monthly rainfall

For the first three harmonics the values of Fourier coefficients (A_1, A_2, A_3, B_1, B_1, B_3) were found to be -48.751, -15.263, 19.351 and -67.92, 49.08, -8.028 respectively. With these coefficients and using Equations (6) the periodic component (P_t) resulting from periodic deterministic process may be mathematically expressed as:

$$P_t = 53.165 - 48.751 \cos\left(\frac{2\pi t}{p}\right) - 67.92 \sin\left(\frac{2\pi t}{p}\right) - 15.263 \cos\left(\frac{4\pi t}{p}\right)$$

$$+49.08 \sin\left(\frac{4\pi t}{p}\right) + 19.351 \cos\left(\frac{6\pi t}{p}\right) - 8.028 \sin\left(\frac{6\pi t}{p}\right) \quad (11)$$

The deterministic cycle component (P_t) was computed by using Equation (11) for all the values of t (t_{max}= 1200). After estimating the periodic component it was removed from, historical time series by subtracting the periodic component from historical time series. This process obtained a new stationery series, S_t, resulting from stochastic non-deterministic process.

Stochastic Component

The presence of stochastic component was already confirmed by plotting the correlogram (Figure 1) of observed series and analysis of serial correlation coefficient (SCC) and coefficient of variance (CV). The periodic component was removed from the historical series. The rest of the data were analyzed to obtain non-deterministic stochastic component by fitting the autoregressive or Markov process of stochastic modeling.

Selection of Model Order

Residual variance method was used to determine the order of the model which may significantly represent the non-deterministic stationary stochastic component. Residual variance at different lags was computed. The minimum residual, variance was obtained for order two. The values of residual variance after 3rd order showed no definite trend.

Using Equation (7) and the estimated autoregression coefficients the stochastic component of the monthly rainfall time series may be expressed as:

$$S_t = 0.456\, S_{t-1} - 0.087\, S_{t-2} + a_t \quad (12)$$

Where S_t, S_{t-1}, S_{t-2} are stochastic component at time t, (t-1), (t-2).

Residual Series of Stochastic Component

The residual series (a_t) which is random independent part of stochastic component was obtained after removing the periodic and dependent stochastic parts from the historical series.

The statistical analysis of the residual series confirms its normal distribution with mean which is almost equal to zero (mean 0.00 and SD 3.81). The values of statistical measures are presented is Table 2. The mean, SD of the historical and generated series are almost same which shows closeness between historical and generated data.

Table 2. Statistical parameters of the observed, generated and residual series of monthly rainfall

Parameters	Historical series	Generated series	Residual series
Mean, mm day^{-1}	53.16	53.16	0.00
SD, mm day^{-1}	91.57	71.58	3.82
Variance	8385.06	5123.75	14.58

Model Structure

Since the observed monthly rainfall series was found to be a trend free series that developed model describes the periodic-stochastic behaviour of the series. The developed model is a superimposition of third harmonic deterministic process and second order autoregressive model. The mathematical structure of the additive model can now be represented as follows:

$$\text{Rainfall} = 53.16 - 48.751\cos\left(\frac{2\pi t}{p}\right) - 67.92\sin\left(\frac{2\pi t}{p}\right) - 15.263\cos\left(\frac{4\pi t}{p}\right) + 49.08\sin\left(\frac{4\pi t}{p}\right)$$

$$+19.351\cos\left(\frac{6\pi t}{p}\right) - 8.028\sin\left(\frac{6\pi t}{p}\right) + 0.456\, S_{t-1} - 0.087\, S_{t-2} + a_t \qquad (13)$$

The first seven terms in the formulated model represented by Equation (13) constitute the deterministic part of the monthly rainfall time series. The eighth and ninth terms represent the dependent stochastic component of the model where the current value of S_t depends on the weighted sum of observed preceding two values. The last term is the random independent part of the stochastic component. Using the developed model the average monthly rainfall series was generated for all the values (t=4 to 1200).

Diagnostic Checking of Rainfall Model

The residuals obtained after fitting the formulated model was subjected to various analysis to test their adequacy for representing the time dependent structure of the monthly rainfall.

Sum of squares analysis

The sum of squares of residuals series were compared with sum of squares of deviations of observed values from their mean.. The value of coefficient of determination (R^2) was found to be 0.9949, which is nearly equal to unity. Thus, this leads to the conclusion that the developed model has a fair goodness of fit to generate the monthly rainfall series.

Serial correlation analysis

The serial correlation coefficients (SCC) for lags l (l= 1, 2, 3....36) were computed with the help of Equation (10). The values of SCC against respective lags were then plotted to obtain a correlogram. The resulting correlogram is shown in Figure 3 with confidence limit at 1 per cent level. The correlogram is almost completely contained within the confidence limits at 1 per cent level. Hence it may be treated to be non significant. Further the correlogram shows that at 1 per cent level the coefficients are within the limits. This again confirms that the residual series may be treated to be non significant.

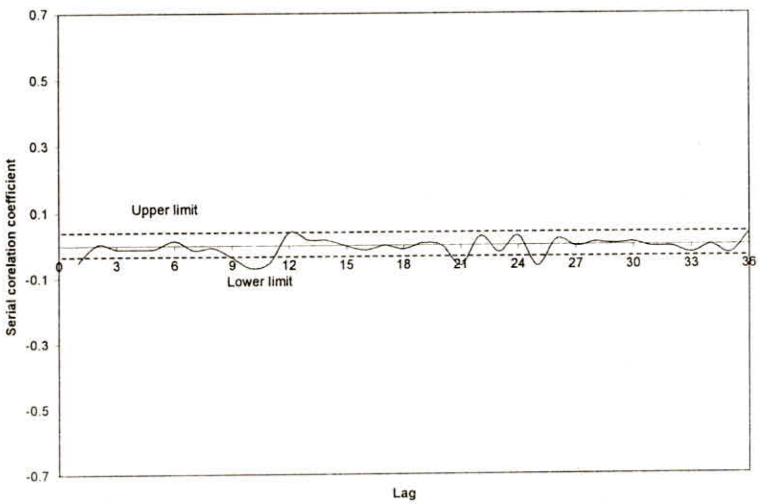

Figure 3. Correlogram up to lag 36 for residual series of monthly rainfall for 100 years (1901-2000) at Udaipur

The residual series has a mean value of 0.00 (zero) and the variance of 14.58 which is approximately equal to 1/150th variance of historical series. This leads to the conclusion that the residuals are independent normally distributed. Further, it also confirms the randomness of the residuals.

Validation of Stochastic Model of Monthly Rainfall

Validation of generated monthly rainfall series by developed stochastic model (Equation 13) was done by comparison of generated monthly rainfall series and measured monthly rainfall series. Validation of generated 100 year mean monthly rainfall series was made with 100 year mean measured rainfall series (Figure 4). The relationship is shown in Figure 5. The correlation coefficient between generated mean monthly rainfall series and measured mean monthly rainfall was found to be 0.9974. The correlation was tested by t test and found to be highly significant at 1 per cent level. The standard error (5.57 mm) is quite low. The mean of the monthly generated rainfall was found to be 53.165 mm. Mean of the measured monthly rainfall series was found to be 53.165. The regression equation is very near to 1:1 line. Therefore, Equation (13) can be used for future prediction of monthly rainfall.

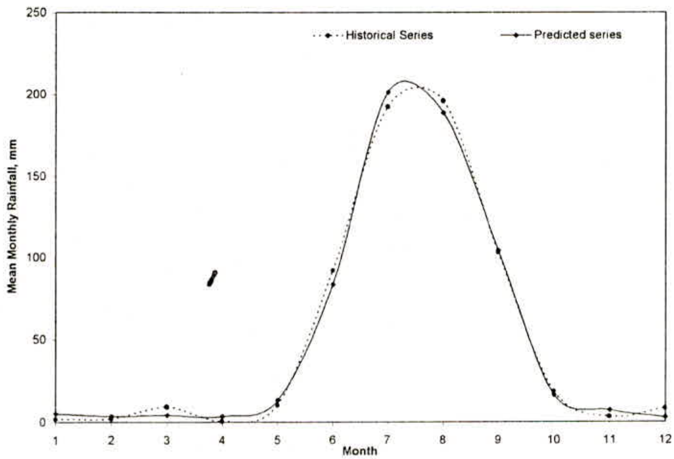

Figure 4 Variation of generated mean monthly rainfall and measured mean monthly rainfall for 100 years (1901-2000) at Udaipur

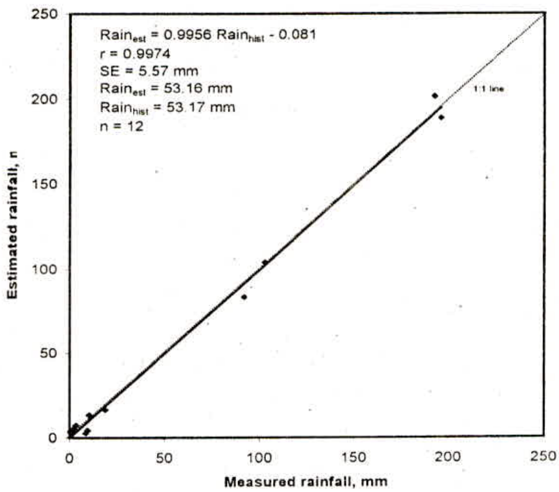

Figure 5. Relationship between generated mean monthly rainfall ($Rain_{est}$) and measured mean monthly rainfall ($Rain_{hist}$) for 100 years (1901-2000) at Udaipur

REFERENCES

Bhakar S. R. (2000), Modelling of evaporation and evapotranspiration under climatic conditions of Udaipur. Ph. D. thesis submitted to faculty of Agricultural Engineering, Department of Soil and Water Engineering, CTAE, MPUAT, Udaipur.

Kottegoda, N.T. (1980), Stochastic water resources technology. The Macmillan Press Ltd., London: pp. 384.