STOCHASTIC MODELLING OF RAINFALL

1.0 INTRODUCTION

For appropriate planning and operation of water resources projects, long term historical records of stream flow are required. Where the historical records of stream flow are short, the generation of synthetic sequences of stream flow provides alternate data series for planning decisions. There are also applications when interest of hydrologists extends beyond the length of the available historical record of the rainfall. Only probabilistic models, applied to existing series of measurements, can provide such an extension. Because of continuous development and short stream flow record, observed stream flow data are generally not representative of current catchment conditions. On the other hand, rainfall records are generally longer than stream flow records and are also generally not affected by the developments in the catchment. Also, for a given length of record the statistical characteristics of rainfall can be defined with relatively less error than in the case of stream flow and hence can be used with stochastic data generation models. The generated rainfall can be used as input to deterministic models for synthesis of stream flow data.

All statistical calculations performed on hydrological data, no matter how sophisticated, will reduce the amount of information included in the original record. A synthesis, which is a result of all statistical operations, is nothing but a useful way to practical aplications. Yevjevich (1972) suggested that hydrologic time series can be modelled by a deterministic component and a stochastic component. The deterministic component is composed of trends, jumps and periodicities whereas the stochastic component is reflected by the randomness of the hydrologic variable (Figure 1a and 1b).

2.0 DATA GENERATION

Data generation techniques or Monte Carlo simulation have been widely used in hydrology. These range from generating large samples of data from known probability distributions to studying the probabilistic behaviour of complex hydrological processes. The value of the model depends on its ability to generate new rainfall series which correctly reproduce in a statistical sense characteristics that are observed in the historical series.

2.1 Stochastic Models

A stochastic model is a probabilistic model having parameters that must be obtained from observed data. Stochastic models contain random components. These random components contain random elements. If a stochastic model is to be used to generate hydrologic data, methods must be available for generating the random elements of the models.

Random element is usually thought of as an element selected in a fashion such that each element in the population has an equal chance of being selected. More generally a random element can be selected from any probability distribution as long as the elements are independent of each other.

2.2 Stochastic Rainfall Models

Models of point rainfall time series have potential application to a range of hydrological problems like the generation of rainfall across a range of time scales for hydrological design and the disaggregation of large time interval data for short duration application.

Rainfall is a natural process which results from the interaction of complex atmospheric processes. Because of the complexity of the process , rainfall cannot be described in purely deterministic terms. The rainfall process also contains periodic components due to the seasonal variation within the year and persistence in both time and space. One way to produce a rainfall pattern sysnthesis consists in the mathematical simulation of rainfall series. Two different approaches are available, either a deterministic or astochastic procedure. A deterministic approach would use hydrometeorological information about atmospherical conditions and known physical laws in order to model situations when rainfall can occur. This approach includes, for example the numerical model developed by Georjakakos (1987). Another example is the model developed by Collier and Hardaker (1996) and Andrieu, et al (1996). Such models contribute to the basic understanding of rainfall generating processes, but can hardly find application in practical hydrology problems because they require data of a large number of geographical, oceanographical and meteorological parameters.

Because of the complexity and strong dependence upon initial conditions of the precipitation process, a stochastic approach is likely to be preferable to a purely physical model. In surface water hydrology, Monte Carlo simulations are applied assuming the rainfall process and certain catchment characteristics to be stochastically varying in space and time. In an effort to provide more concise models of daily rainfall, several investigators have proposed stochastic models describing both rainfall occurrence and the distribution of rainfall amounts at a point in space.

In the beginning of the twentieth century the study of stochastic structure of time series of rainfall data began. It was observed that the wet and dry weather sequences have persistence. Mathematical rainfall simulation models have traditionally used a stochastic approach for generation of space time rainfall fields. This is partly due to the lack of knowledge regarding the physical mechanisms which govern the spatial and temporal variability of rainfall and partly due to the difficulties in finding sufficiently detailed spatial rainfall data with the desired temporal resolution. The stochastic models constitute an important step forward as regards the parameterisation of the spatial and temporal character of the rainfall fields. The statistical properties of the observed rainfall can be used in the generation of future events. Mathematical models which can produce rainfall series are called stochastic models.

Several stochastic models have been developed during the last two decades for the daily rainfall occurrence. Early studies assumed that each storm was made up of a random number of rain cells which occur in space and time according to a three dimensional point process and the resulting rainfall could be described by a Poisson process. Later studies, however, considered rainfall to occur in clusters and used the Neyman-Scott model in place of Poisson models. Cluster centers in the Neyman Scott model are not rain cells, but are just points around which the density of cells is larger than in other regions. Each cluster has associated with it a number of rain cells which is a random variable, independent and identically distributed for each cluster center. In the cluster process the number of cells in a storm is randomly distributed and the cell arrival times are exponentially distributed.

Recently, the use of Poisson cluster process in stochastic modelling of rainfall has been investigated by several authors. The cell arrivals are modelled by a Poisson cluster process i.e. storm arrivals form a Poisson process and a cell arrival distribution is assigned to each storm; the depth and duration of the cell are modelled by exponential distributions.

The methodology used for generation of daily rainfall consists of two parts : the first determines the occurrence of dry and wet days and the second generates the rainfall depth on wet days. For reproducing the occurrence of the rain events, the techniques as discussed in the following sections have been used.

2.2.1 Markov Chain

Many hydrologic time series exhibit significant serial correlation. That is, the value of the random variable under consideration at one time period is correlated with the values or the random variable at earlier time periods. The correlation of a random variable x at one time period with its value k time periods earlier is denoted by $\rho_x(k)$ and is called the kth order serial correlation. If $\rho_x(k)$ can be approximated by $\rho_x(k) = \rho^k_x(1)$, then the time series of the random variable x might be modelled by a first order Markov process. A first order Markov process is defined by the equation:

$$X_{i+1} = \mu_x + \rho_x(1)(X_i - \mu_x) + \epsilon_{i+1} \tag{1}$$

where, X_i is the value of the process at time i , μ_x is the mean of X, $\rho_x(1)$ is the first order serial correlation and ϵ_{i+1} is a random component with $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2$. This model states that the value of X in one time period is dependant only on the value of X in the preceding time period plus a random component. Further it is also assumed that ϵ_{i+1} is independent of X_i. If the distribution of X is $N(\mu_x, \sigma_x^2)$ then the distribution of ϵ is $N(0, \sigma_\epsilon^2)$. random value of X_{i+1} can now be genrated by selecting ϵ_{i+1} randomly from a $N(0, \sigma_\epsilon^2)$ distribution.

Thus a model for generating X s that are $N(\mu_x, \sigma_x^2)$ and follow the first order markov model is:

$$X_{i+1} = \mu_x + \rho_x(1)(X_i - \mu_x) + t_{i+1}\sigma_x\sqrt{1-\rho_x^2(1)} \tag{2}$$

The procedure for generating a value for X_{i+1} is to estimate μ_x , σ_x and by X, S_x and $r_x(1)$ respectively and then select t_{i+1} at random from a N (0, 1) distribution and calculate X_{i+1} based on X, S_x and $r_x(1)$ and X_i In this approach events are considered to belong to a certain number of states. No rain is one state and the other states which are wet may be only one or several. The probability of a day belonging to a certain state is dependent on the occurrence of belonging to a certain state is dependent on the occurrence of several previous states.

Equation (2) generates normally distributed X s with a mean of μ_x variance of σ_x^2 and first order serial correlation of $\rho_x(1)$. For a first order Markov process, the lag k serial correlation $\rho_x(k)$ is given by:

$$\rho_x(k) = \rho_x^k(1) \tag{3}$$

Thus, the correlogram exponentially decays from $\rho_x(0) = 1$ to $\rho_x(\infty) = 0$ according to equation (3). If an oserved correlogram has this property, the Markov model may be an appropriate generating model.

Haan et. al. (1976) developed a stochastic model based on a first order Markov Chain to simulate daily rainfall at a point. The model uses historical rainfall data to estimate the Markov transitional probabilities. The model was said to be capable of simulating daily rainfall records of any length based on the estimated transitional probabilities and the frequency distribution of rainfall amounts.

Roldan and Woolhiser (1982) and Woolhiser and Roldan (1982) used a first order Markov chain as the occurrence process and a mixed exponential distribution for the daily rainfall. The model was reported to have performed better than other alternatives.

2.2.2 Transition Probability Matrix Method

The essential features of this method are:

(i) the range over which rainfall is expected to vary is divided into set of discrete intervals.

(ii) a matrix corresponding to the intervals is built up from the observed rainfall sequences by tabulating the number of times the observed data went from state i to state j denoted by n_{ij} as follows :

Final state

Starting state

	0	1	2	3
0	n_{00}	n_{01}	n_{02}	n_{03}
1	n_{10}	n_{11}	n_{12}	n_{13}
2	n_{20}	n_{21}	n_{22}	n_{23}
3	n_{30}	n_{31}	n_{32}	n_{33}

(iii) the m x m transitional probabilities matrix is $p = [p_{ij}]$ which is given by

Final state

Starting state

	0	1	2	3
0	p_{00}	p_{01}	p_{02}	p_{03}
1	p_{10}	p_{11}	p_{12}	p_{13}
2	p_{20}	p_{21}	p_{22}	p_{23}
3	p_{30}	p_{31}	p_{32}	p_{33}

where, $p_{00} = n_{00} / (n_{00} + n_{01} + n_{02} + \ldots\ldots)$ (4)

$p_{10} = n_{10} / (n_{10} + n_{11} + n_{12} + \ldots\ldots)$ (5)

and so on each p_{ij} being obtained by dividing n_{ij} by the corresponding row total of the n_{ij}.

(iv) After the transitional probabilities are estimated the next step is the simulation of rainfall using appropriate distribution. The synthetic sequences were generated by dividing the rainfall into a number of classes (intervals). The sequence of states are built up by selecting a pseudo random number between 0 and 1 and then assigning the state according as the value of u is less than or greater than $p_{01} + p_{00}$ and then moving to the next state and so on

After trying as many as 13 classes Haan et. al. (1976) found six classes to be a reasonable choice. Srikanthan and Mc Mohan (1982) used seven classes. The rainfall in the last class was generated using a shifted exponential distribution. Haan et. al. (1976) used a multi state 7 x 7 Markov chain model and employed a uniform distribution for each of the wet states except for the last for which an exponential distribution was assumed. The model was tested on data of seven rainfall stations in Kentucky. A separate transition probability matrix was used for each month. The class distribution for the states in the Markov chain were found by using geometric progression. The comparison of simulated rainfalls with observed rainfalls indicated that the model generated rainfalls exceeded the historical rainfalls on an average by about 2.5%.

2.2.3 Alternating Renewal Process

This process consists of alternating wet and dry spells. Wet spells are assumed independent and belong to a particular distribution. Similarly dry spells are assumed independent and belong to another distribution. Further, the two random sequences independent. The following definitions apply :

(i) A wet spell is a sequence of wet days bounded on either side by a dry day.
(ii) A dry spell is defined likewise
(iii) Spells are assigned to the periods (usually months or seasons) in which they begin.
(iv) A day is defined as wet if the rainfall exceeds a threshold value δ mm

The first two assumptions could be satisfied by analysisng data on a monthly or seasonal basis. The third assumption is checked by computing the correlation between wet and dry spells in each month.

Several distributions can be fitted to the data to model the lengths of wet and dry spells. Commonly used distributions for wet and dry runs are ; the truncated negative bionomial distribution (TNBD) and the shifted negative bionomial distribution (SNBD). The probability density function of the TNBD is given by:

$$P(x=K|x\geq 1) = \left[\frac{K+r-1}{K}\right] p^{r} \frac{(1-p)^{r}}{1-p^{r}} \qquad (6)$$

in which x is the random variable, k is the length of the spell and p and r are parameters ($0 < p < 1$; $-1 < r$). If the rainfall is modelled on monthly basis, 24 parameters are required to be estimated for equaton (6) for the whole year. In general, the development of rainfall model based on the alternating renewal process requires long data series so that sufficient number of wet and dry spells could be included. Data of 25 to 30 years when used for only a season are known to have performed well.

The model for rainfall amounts must be one which describes the distribution of rainfall amounts on days when it rains. The distribution is highly skewed and since the wet day is defined as a day on which rainfall exceeds a threshold value δ mm, a shifted or truncated distribution would model the rainfall amounts well.

The shifted two parameter gamma distribution (SGD) has been frequently used to fit the rainfall amounts. The probability density function of this distribution is given by:

$$f(y) = \frac{\lambda^{\nu} y^{\nu} \exp(-y\lambda)}{\Gamma \nu} \qquad (7)$$

in which y is the rainfall amount, Γ is the gamma function, and ν and λ are parameters.

The mean rainfall of a wet spell depends on the length of the spell (Buishand, 1978). To account for this, three types of rainfall are distinguished. These are :

(i) wet spells with a solitary wet day (type 0)
(ii) a wet day with one adjacent day also wet (type 1)
(iii) a wet day with both adjacent days wet (type 2)

The probability density function in Eq. (7) is derived separately for each type of wet spell for each month. The estimation of parameters p and r in Eq. (6) and ν and λ in Eq. (7) is done using the method of maximum likelihood.

The generation of daily rainfall sequences is carried out in two steps. First, the lengths of wet spells and dry spells are generated by coupling a uniform random number in the interval (0,1) to the cumulative distribution function of TNBD. For wet spells of type 0, type 1 or type 2 the appropriate random gamma variates are generated to obtain the values of daily rainfall.

2. 3 Studies in India

In India, Seth and Obeysekera (1979) used the transition probability matrix method for generation of daily rainfall data in the Naula catchment of Ramganga basin in Uttar Pradesh. Singh and Kripalani (1982) studied the dependence in daily rainfall of 12 stations and in daily rainfall of 10 meteorological regions during the summer monsoon by analysing the rainfall as stochastic point process by fitting several types of models like log model, Markov chains of order 1 and 2 to station data.

Rao and Reeves (1991) used the alternating renewal process model for simulation of daily rainfall during monsoon season for one station in western India. They used the TNBD in their study. The authors mentioned that in addition to simulating the monthly means and variances of the historical series well the model also predicted the distributions of the west and dry spells of different lengths adequately. Ramasastri (1993) applied the transition probability matrix and alternating renewal process models for generating daily rainfall sequences using data of historical rainfall series of Jaipur in Rajasthan.

3. 0 GENERATION OF DAILY RAINFALL DATA - CASE STUDY

Generation of daily rainfall data has been done for one station Jaipur in east Rajasthan. Using 30 years (1961 - 1990) of historical rainfall data, ninety years of data are generated by the Transition Probability Matrix Approach. The statistical parameters used for comparison of historical and simulated series are:

(i) mean and standard deviation of monthly rainfall
(ii) average number of wet days (daily rainfall > 1. 0 mm)
(iii) magnitude of maximum daily rainfall during the 30 years historical and simulated rainfall

The comparison was done only for the monsoon season since rainfall during the remaining part of the year is not significant.

For generation of daily rainfall data with the Transition Probability Matrix (TPM) method , seven classes of rainfall were used. After a few trials with different values for the class limits, the following class boundaries have been decided.

Table 1 : Class Boundaries (Trial I)

	Class in	Boundary mm
Class Number	Lower	Upper
1	0. 001	0. 9
2	1. 0	2. 9
3	3. 0	6. 9
4	7. 0	14. 9
5	15. 0	30. 9
6	31. 0	62. 9
7	63. 0	∞

Depending on the highest observed rainfall and the average number of rainy days in each month, varying number of classes have been used for different months. If the number of classes is k , the upper class boundary for the kth class will be infinity. All the class boundaries given in Table 1 above, therefore, apply only to a month where all the seven classes have been used. The classes used in different months are given below:

Mon	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
class	4	4	3	3	5	6	7	7	6	4	4	4

Using the TPM model with the Box Cox Transformation for the last class and the apropriate number of classes for different months as above, ninety years of data has been generated. The generated data has been split into three samples of 30 years each and the different parameters for comparison with the historical series have been computed and averaged. The results are presented in Table 2 (a) to Table 2 (c). The mean and standard deviation of the monthly rainfall computed from the generated daily rainfall data compared well with those of the observed series. The mean number of wet days (days of daily rainfall more than 1.0 mm)

Table 2(a) : Mean and Standard Deviation of Historical and Simulated Rainfall for Monsoon Months - Station: Jaipur

Method : TPM Trial I Units : mm

Month	Historical		Simulated (I)		Simulated (II)		Simulated (III)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
June	58.6	74.0	81.4	66.1	49.1	31.9	62.3	54.3
July	227.1	181.4	250.6	125.8	198.3	122.0	175.4	115.3
August	209.3	74.6	221.1	72.3	221.0	98.0	213.5	94.4
Sept.	74.6	74.2	69.3	39.7	70.6	49.6	92.0	55.6

Table 2 (b) : Average Number of Wet Days* for Monsoon Months

Method : TPM Trial I Station : Jaipur

Month	Historical	Simulated (I)	Simulated (II)	Simulated (III)
June	4.4	5.0	5.0	4.4
July	12.4	12.3	11.7	11.7
August	13.0	13.4	13.9	13.9
September	5.6	5.9	5.7	6.5
Season	35.4	36.6	36.3	36.5

* A day with rainfall of 1.0 mm or more

Table 2(c) : Magnitude of Maximum Daily Rainfall During 30 years

Method : TPM Trial I Station : Jaipur Units : mm

Historical	Simulated (I)	Simulated (II)	Simulated (III)
287.0 in July	130.4 in June	168.0 in July	160.8 in July

To generate the extreme rainfall matching with the observed maximum one day rainfall another trial has been made by changing the limits of the last class and making modifications to retain seven classes. The limits of the seven classes have been taken as given in Table 3.

Table 3 : Class Boundaries (Trial II)

	Class	Boundary mm
Class Number	Lower	Upper
1	0. 001	0. 9
2	1. 0	6. 9
3	7. 0	14. 9
4	15. 0	30. 9
5	31. 0	62. 9
6	63. 0	124. 9
7	125. 0	∞

The 90 years generated data have been divided into three samples of 30 years each as in the case of Trail I. The results of comparison are presented in Table 4 (a) to Table 4 (c). It may be seen that the generated data with class boundaries as in trail II was able to generate extreme rainfall values comparable in magnitude to the observed extreme one day rainfall at Jaipur in July 1981. Also, the mean and Standard deviation and the number of wet days in the different months of monsoon season were comparable with the corresponding statistical parameters of the historical data.

Table 4 (a) : Mean and Standard Deviation of Historical and Simulated Rainfall for Monsoon Months - Station : Jaipur

Method: TPM Trial II Units : mm

Month	Historical		Simulated (I)		Simulated (I)		Simulated (I)	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
June	58.6	74.0	73.7	70.7	80.7	65.4	74.5	57.1
July	227.1	181.4	224.6	146.9	265.9	168.2	230.3	134.7
August	209.3	74.6	221.9	118.5	219.5	79.8	192.9	78.2
September	74.6	74.2	77.7	59.7	78.9	60.4	78.1	60.2

Table 4 (b) : Average Number of Wet Days* for Monsoon Months

Method: TPM Trial II Station : Jaipur

Month	Historical	Simulated (I)	Simulated (II)	Simulated (III)
June	4.4	5.1	4.9	5.1
July	12.4	12.7	12.2	12.5
August	13.0	12.6	13.7	13.4
September	5.6	6.3	6.5	6.2
Season	35.4	36.7	37.3	37.2

* A day with rainfall of 1.0 mm or more

Table 4 (c) : Magnitude of Maximum Daily Rainfall During 30 Years

Method: TPM Trial II Station : Jaipur Units : mm

Historical	Simulated (I)	Simulated (II)	Simulated (III)
287.0	255.0	252.7	277.8

4.0 CONCLUDING REMARKS

The performance of a stochastic model is judged to be satisfactory if the model parameters preserve as much of the rainfall characteristics as possible. In any application of data generation methods, it must be kept in mind that data generation cannot improve or overcome faulty data. At best one can generate a set of data having statistical properties equal to the properties of the sample used in estimating the population parameters. In addition to this, data stochastically generated is subject to the same sampling errors as natural data.

REFERENCES

1. Andrieu, H, M. N. French, V. Thauvin and W.F. Krajewski, 1996. *Adaptation and application of a quantitative rainfall forecasting model in a mountainous region.* Jour. of Hydrology, Vol. 184, No. 3 - 4, pp. 243 - 260.

2. Buishand, T.A., 1978. *Some Remarks on the use of daily rainfall models.* Jour. of Hydrology, Vol. 36, pp. 295-308.

3. Collier, C.G. and P.J. Hardaker, 1996. *Estimating PMP using a storm model approach.* Jour. of Hydrology Vol. 183, No. 3 - 4, pp. 277 - 306.

4. Georgakakos, K.P. and M.L. Kavvas, 1987. *Precipitation analysis, Modelling and precipitation in Hydrology.* Review of Geophysical Research, Vol. 25, pp. 163 - 178.

5. Haan, C.T. , D. M. Allen and J.O. Street, 1976. *A Markov chain model of Daily Rainfall.* Water Resources, Vol. 12, No. 3, pp. 443 - 449.

6. Ramasastri, K.S., 1993. *Space Time Characteristics of Rainfall in East Rajasthan.* Ph.D. Thesis, Univ. of Roorkee, Roorkee.

7. Rao, N.H. and D.H. Rees, 1991. *Simulation of daily rainfall in the monsoon season.* Jour. of Indian Water Resou. Society, Vol. 11, No. 3, pp. 23 - 27.

8. Roldan, J. and D.A. Woolhiser, 1982. *Stochastic daily precipitation models (1). A comparison of occurrence processes.* Water Resources Research, Vol. 18 No. 5, pp.1451-1461.

9. Seth, S.M. and J.T.B. Obey Sekera, 1979. *Multi station daily rainfall generation.* Proc. of International Conf. on Computer Applications in Civil Engineering, Roorkee (India).

10. Singh S.V. and Kripalani, 1982. *Some Characteristics of Daily Summer Monsoon Rainfall over India.* Proc. of International Symp. on Hydrometeorology, AWRA, pp. 589 - 596.

11. Srikanthan and McMohan, 1982. *Stochastic Generation of Daily Rainfall of Twelve Australian Stations.* Agricultural Engg. Rept. 57/87, Deptt. of Civil Engineering, University of Melbourne. Australia.

12. Woolhiser, D.A. and J. Roldan, 1982. *Stochastic daily precipitation models (2). A comparison of distribution.*

13 Yevjevich, V., 1972. *Structural analysis of hydrologic time series.* Hydrology Paper 56, Colorado State University, Fort Collins, Colorado, USA.

:: :::: ::

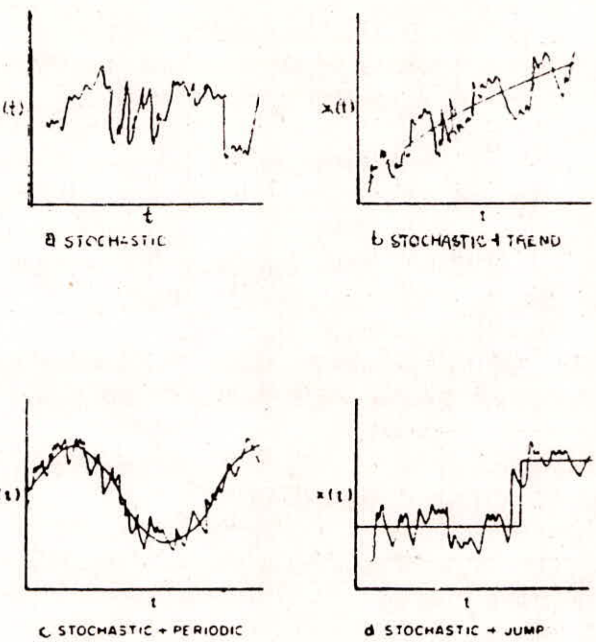

FIGURE 1. A TIME SERIES CONTAINING STOCHASTIC AND SEVERAL TYPES OF DETERMINISTIC COMPONENTS

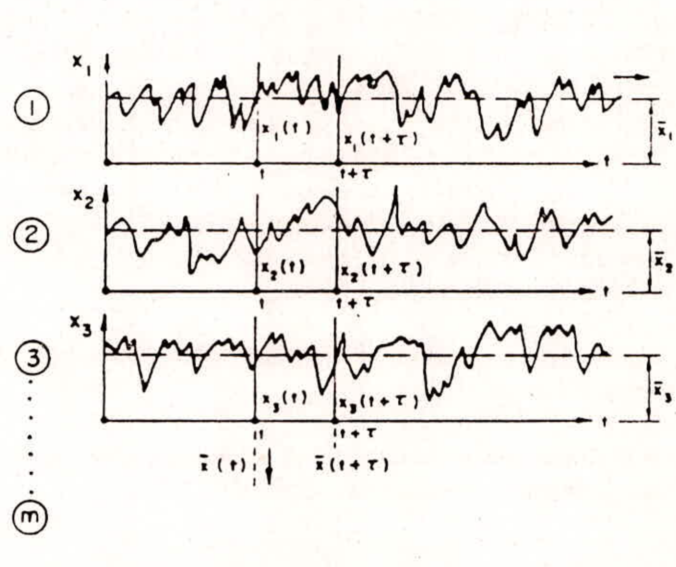

FIGURE 1. B SEVERAL REALIZATIONS OF A STOCHASTIC PROCESS.