SA-1

HYDROLOGIC TIME SERIES MODELLING - AN OVERVIEW

ARUN KUMAR
ASSOCIATE PROFESSOR
DELHI COLLEGE OF ENGINEERING
NEW DELHI

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAVAN
ROORKEE-247667 (UP)
INDIA

1982-83

PREFACE

The first meeting of working group of priority area of 'hydrologic analysis of stream flows in a basin'was held on 16.4.1982 . The members attending the meeting felt that there is a need for preparation of a state-of-art report in this area highlighting the development in India as well as other countries so that priorities of research work in this area can be defined.

Dr Arun Kumar, Associate Professor, Delhi College of Engineering, Delhi was invited to National Institute of Hydrology during 8-28 June 1982 for preparation of the state-of-art report on 'Hydrologic Time Series Modelling- an Overview' in discussions with the scientists in NIH.

CONTENTS

ABSTRACT

Hydrologic variables might vary in space, as well as time and are represented by either continuous or discrete series. Much of the statistical methodology is concerned with models in which the observations are assumed to vary independently. However, a great deal of the data in water resources planning and management occur in the form of time series where observations are dependent and the nature of this dependence is an important characteristic of time series.

The report gives a short review on hydrologic time series with particular emphasis on river flow time series modelling. The analysis of deterministic and stochastic components of time series has been discussed in the beginning and further extended to stochastic models. Short memory models like autoregressive models, moving average models, autoregressive moving average models, autoregressive integrated moving average models have been discussed in the light of identification of the model, parameter estimation and diagnostic checking. Long memory models like fractional Gaussian noise models , fast fractional Gaussian noise models and broken line models have also been described. Details of multisite short memory and long memory models have been provided. Final remarks including areas in which further research is needed and extensive list of references is given at the end of the report.

1.0 INTRODUCTION

In recent years, stochastic modelling and time series analysis have gained tremendous importance in hydrological studies. This is because primary hydrological variables like rainfall, lake levels, riverflows etc. are essentially stochastic in nature and are not amenable to solution by the classical statistical methods. Since the number of research papers in time series modelling have been phenomenal in recent years, it is impossible to encompass all the details of modelling in a short review paper. The aim of this study is, therefore, limited to bring forward to the practising hydrologists an overview on general broad areas of stochastic modelling. The present study mainly concentrates on the stochastic behaviour of riverflow time series and discusses its modelling and applications by hydrologists.

A time series is a continuous set of observations that measures some aspect of the phenomenon, for example, the discharge or stage, in a river with respect to time. The time series may be continuous or discrete. It is a continuous time series, if the hydrologic variable is measured continuously with reference to time, for example, the chart of an automatic flow gauge on a river. In discrete time series, the continuous hydrologic variable is measured in discrete intervals of time for example, the hourly measurement of stages. Generally the time series analysis is done on discrete series because, unlike a continuous series, it can easily be handled by a digital computer. The hydrologic time series is stochastic in character as the output of a hydrologic event from a given input is not unique. For example, the discharge in a river is, primarily, a function of the effective rainfall during the previous period. This functional transfer relationship between the effective rainfall and runoff is dependent on many interrelated physiographic and meteorological factors which cannot be quantified explicitly, making the riverflow stochastic.

The main objective of studying hydrologic time series is to understand the mechanism that generates the data so that the future sequences may be simulated or to forecast the future events over a short period of time(forecasting). These are attempted by making inferences regarding the underlying laws of the stochastic process from the historical data and then by postulating a model that fits the data which can in turn be used for simulating/forecasting the future values. It is, therefore, necessary to identify the various components of the hydrologic time series.

1.1 Hydrologic Time Series

In general, a time series can be divided into two components, viz., deterministic and stochastic. Deterministic component is the one which can be determined by the predictive means, whereas, the stochastic component consists of chance and chance dependent events. Hydrologic time series has both these components. The deterministic components are in terms of trend (increasing or decreasing flows as the time increases) and cyclicities (over the years or within year) and the stochastic component due to erratic atmospheric circulation. The annual cyclicities in the hydrologic process are produced by the annual astronomical cycle e.g. in a given river basin high intensity and high frequency of precipitation in wet season and low intensity and low frequency of precipitation in dry season are expected. The low runoff in the dry season results mainly from the groundwater effluence with relatively small variation, while the high runoff in wet season is formed either by highly fluctuating rainfall or snowmelt or both. Thus the mean and the variance of the stream runoff are large in wet season and small in dry season. This phenomenon indicates within the year periodicity in mean and variance. Usually periodicity in a hydrologic time series would appear in such statistics like means, standard deviation, etc. However, the periodicity is deterministic and can be identified in a given series.

After the deterministic components like the trend and the periodicity are removed from the data, the residual series is the stochastic series. The stochastic nature of the hydrologic time series is caused by such factors as the variable opacity of the atmosphere to solar radiation, fluctuating turbulence, large scale vorticities and heat transfer in the atmospheric, oceanic and continental air and water movements. However, there may exist certain amount of time dependence in the stochastic series which may be created or increased by water storage of various types in the hydrologic environment. A stochastic model may be required to model this time dependence.

The analysis of hydrologic time series is to identify the deterministic i.e. the trend, the periodicity components in the series and separate them from the original series. The resultant stochastic series is then modelled and the combined effect is determined by superimposing them. An excellent discussion on determinism and stochasticity in hydrological time series is given by Yevjevich (1974).

2.0 DETERMINISTIC COMPONENT ANALYSIS

2.1 Trend

A steady and regular movement in a time series, through which the values are, on the average increasing or decreasing, is termed as trend. The existence of trend in hydrological series may be due to low frequency oscillatory movement induced by climatic changes or through changes in land use and catchment characteristics. Many hydrologists have the view that hydrological (river flow) time series have no important trends which can be identified by statistical analysis since the typical length of the series being less than 50 years (trend analysis is generally done on annual series and not on seasonal series so as to supress the effect of periodic component) cannot reflect the long term climatic changes. It is, however, quite likely that there may be an adhoc change in the mean flow in a river due to some abstraction of water from one river to another or because of construction of some reservoirs. In such cases, the trend analysis is generally limited to adhoc modification in the mean. Such a study was done by Smirnov (1969) for the flow of the Volga river at Volgograd.

However, if a trend in a particular series is obvious, it can be described by fitting a polynomial equation of the form given in Eq.(1) to the original x series.

$$X_t = x_o + \alpha_1 t + \alpha_2 t^2 + \text{--------} + \alpha_n t^n + Z_t \qquad \ldots(1)$$

Where $\alpha_1, \alpha_2, \alpha_n$ are estimated by least square.

Generally, a simple linear type model is sufficient

$$X_t = x_o + \alpha t + Z_t$$

where Z_t is the residual series

Many a times, the presence of trend in a series is not obvious. For this reason, a number of statistical tests for detecting trend have been devised. If a series is thought to have a trend component, Kendall's rank correlation test (Kendall and Stuart, 1973) can be used to test its significance. This test, referred to as τ-test determine whether the series is trend free or not. It may, however, be mentioned that rank correlation test as well as the linear regression test are not valid for detecting, the presence of non-homogenities like 'Jumps' in the series. The important test for detecting the presence of jumps in the series is given by Buishand (1977) using Von Neumann's ratio method.

Many a times, smoothing techniques are first used before the trend analysis is attempted. Smoothing techniques enables to bring certain systematic behaviour in the observed series. Smoothing techniques were first used by communication engineers in their attempt to separate signals from noise. Of the various smoothing techniques a linear moving average model is the most generally used. Durbin (1962) has given mathematical justification to these techniques. An undesirable consequence of this type of trend removal is that artificial cycles may be induced into the data. This is known as 'Slutzky-Yule' effect (1937). To circumvent this problem, harmonic and other weighted type of trend removal has been applied in meteorology (Holloway, 1958, Brier, 1961). Smoothing techniques have directly been borrowed from communication engineering literature. This may be quite useful to separate signals from noise, but are quite laborious if used in natural time series. Generally, these methods should be used in conjunction with spectral analysis. The details of various smoothing techniques are given by Kendall and Stuart (1976) and Brown (1963).

It may be observed from the references that hydrologists have not contributed significantly towards trend analysis. Absence of trend means that the time series statistical characteristics like mean, standard deviation, skewness etc are remaining same over the time space. Lack of discussion on trend analysis in hydrologic literature reflect that stationarity (absence of trend) is generally taken as an implicit aspect of hydrologic modelling. A danger in using trend techniques is that low frequency random effects may be taken as trends and this in turn may drastically affect the correct modelling of low frequency values. Since the length of hydrologic record is of very short duration, one hydrological approach is to treat trend like behaviour (which cannot be identified with changes in catchment characteristics) as low frequency movement; hydrologic models which cope with low frequency effects are then used directly.

2.2 Periodic Component

2.2.1 Long range periodicity

A hydrologist is interested to know whether the series has long range periodicity (more than a year). For example, it is commonly said that drought occurs once in five years. The statement implies that drought has a periodicity of five years. Many climatologist and hydrologists in an effort to make long range forecasting have tried to relate the river flow series to various geophysical factors. The phenomenon which received maximum attention from researchers is the sunspots and its likely relationship to precipitation and runoff series. The sunspots number vary approximately in a long term periodic manner, the period ranges from 13 years to 8 years with a mean of 11.1 years. Dixey (1964) showed that there exist a meaningful correlation between the sunspot number and the levels in lake

Victoria. Similarly, Smirnov (1969) found significant correlation between sunspot and the mean flows in the Volga. Indian scientists at Poona have recently correlated changes in the monsoon activity with the sunspot numbers. However, Rodriguez and Yevjevich (1967) investigated the relationship of 88 series of monthly precipitation, 174 series of annual precipitation of U.S.A. and 16 series of annual runoff from all over the world with the sunspot number and they could not find any significant correlation. It, therefore, seems reasonable to assume that the random influences on the hydrological cycle outweighs any effect due to geophysical effects like sunspots. It can be assumed that the observed riverflow series does not follow any long range periodic behaviour.

2.2.2 Short range periodicity (seasonality)

The riverflow series may be obtained as an annual , monthly, ten daily, pented (5 daily) or as a daily observed data. Though the annual data does not follow any long term periodic behaviour, the seasonal cyclic effects are present in other series. Within the year periodicity is due to annual revolution of the earth around the sun, by the moon and by daily rotation of the earth. These are called seasonal effects as they are repeated at the same time in each year and are thus deterministic. They show themselves in monthly data, say, the monthly means and variances being unequal. Similarly, the statistical distribution for different months will be different . Seasonality is also observed in pentad and daily data (Bernier, 1970).

The hydrologist approach has been to identify the seasonality in a given series and removing its component from the original observed series. This is called 'prewhitening' of the series. The resulting series is said to be 'deseasonalised' and assumed to be stationary. It is then modelled by a stationary stochastic model which is finally 'dewhitened' into a seasonal model.

A common method of 'prewhitening' is to 'standardize and remove periodicity from, say an observed monthly sequence where $x_{t,\tau}$,t=1.2....; and τ=1,2...12 are the calender months, by using the formula

$$Z_{t,\tau} = \frac{x_{t,\tau} - m_\tau}{\sigma_\tau} \qquad \ldots . \quad (2)$$

Where m_τ and σ_τ are the mean and the standard deviation of the months. The mean m_τ and σ_τ will have 12 values each in a monthly series. Similarly, in the case of daily flow, m_τ and σ_τ will have 365 values each. In case of short samples, the estimate of such large values may lead to sampling errors. In such cases, the values of m_τ and σ_τ are smoothened by using harmonic analysis. It is observed that the periodicity can be represented by one or two harmonics in monthly series and by four or six harmonics in daily series. Objectively, the actual number of harmonics to be fitted in each case can be found through variance analysis using F-distribution. Extensive work has been done in harmonic smoothing on monthly and daily flows by Roesner and Yevjevich (1966), Quimpo (1967) and Yevjevich (1972). Many times, before standardization, prior transformation of flow data is made, such as, logarithms, in order to reduce the variance effect and to make the distribution more Gaussian. It may, however, be noted that there may still be seasonality in the series such as in correlation and skewness coefficients which is not removed by pre-whitening.

3.0 STOCHASTIC MODELLING

The stochastic series is obtained after the deterministic component has been identified and removed from the original series. Any mathematical modelling will now require the modelling of the stochastic component. If the process is entirely random, then the set of values (Z_1, Z_2 ... ... Z_n), obtained after eliminating the trend and the periodic component will be independent of each other. They then constitute a pure random process. In many hydrologic series, the values of Z_t may depend upon the antecedent values of Z. In other words, there may exist an internal dependence in the time series which may now be written as:

$$Z_t = f(Z_{t-1}, Z_{t-2}, \ldots\ldots) + a_t \qquad \ldots (3)$$

where a_t is the random component. The modelling of the time series, Z_t, depends upon the statistical properties of the Z_t series, the probability distribution to which this series belongs and the structure of the internal dependence. These are briefly explained as follows:

3.1 Probability Distribution

Every series has a particular probability distribution and in order to correctly model a given series, it is of importance to identify the probability distribution to which the series belong. A wide variety of probability distributions have been used to fit hydrologic data and generalisation is not possible. However, it can be said that the annual data are likely to be nearer Gaussian than monthly data and that monthly data will be nearer Gaussian than the pentad data and so on. Departure from the Gaussian distribution means that the flow series have certain skewness which has a direct effect on the river flow properties. For example, it has been observed that the riverflow series are generally, positively skewed. The presence of high skewness means there exist an unusually high flow in the midst of near average flows. Such a situation which is commonly observed in hydrological series is termed as Noah effect (Mandelbrot and Wallis, 1968). Many a time, there is an unusually long period of very low or high flows which is called as Joseph effect. Though Joseph effect can be modelled by the proper selection of the stochastic model, the Noah effect can be correctly modelled only by the proper use of the probability distribution. The identification of the distribution depends upon the coefficient of skewness and kurtosis observed in the sample. It is observed that pentad and daily flows are highly skewed. Seasonal data (after removal of seasonality) generally indicate variable skewness and different family of distributions for different months of the year. The problem of distribution identification gets aggravated because of the sampling errors in the estimation of skewness from the historical data.

In the case of annual flow series modelling the probability distribution can be taken as Gaussian. This was corroborated by the work of Markovic (1965) who fitted various distributions to 2500 annual river-flow and precipitation series across the USA and showed that 75 percent of the annual river-flow series and 90 percent of the annual precipitation series could be fitted by Gaussian distribution at 5 percent significance level. However, in the case of pentad and daily data, family of Pearson distribution need be fitted. There is wealth of information available on fitting distributions to hydrological series, notable being Moran (1957, 70),Domokos (1970), Matalas and Wallis (1973), Singh and Lonquist (1974) and many others.

Fitting a known distribution to the data is based on the apriori assumption that the series follows a particular distribution. However, the observed riverflow series may not follow the specified distribution and such a distribution may therefore yield a poor fit to the data. Alternatively, the data could be reconstituted by some suitable transformation such that the transformed data follows a particular distribution. The most commonly used transformation is the logarithmic transformation which assumes that the logarithms of the original flows follow a normal distribution. Presently such transformations are also assumed apriori without recourse to the analysis of the data and therefore suffers from the same defect as choosing a particular distribution. Chander et al (1978) have suggested the use of Box-Cox transformation which for a particular value of λ can render any given series nearly Gaussian. The advantage of power transformation is that it makes the analysis free from likely errors arising out of incorrect distribution identification.

3.2 Dependence

It is observed that many hydrological series have internal dependence i.e. there is a tendency for a low flow to be followed by a low flow and a high flow is likely to be followed by high flows. If there is a marked persistence in the sequence of flows, there is a notable tendency of both Z_i and Z_{i+1} to be greater than $\bar{Z}$ or both to be less than $\bar{Z}$. Thus there is a distinct tendency for the product $(Z_i-\bar{Z})(Z_{i+1}-\bar{Z})$ to be positive, since it is frequently the product of two terms of the same sign. In hydrological literature dependence has been classified as short term dependence and long term dependence. Whether a series is dependent or not is checked by the following analysis.

3.2.1 Correlation analysis

The sequential dependence in a time series can be determined by autocorrelation analysis. The correlation coefficient ρ_k between Z_t and Z_{t+k} is given by

$$\rho_k = \frac{E\{(Z_t-\bar{Z})(Z_{t+k}-\bar{Z})\}}{\sigma_Z^2} \qquad \ldots(4)$$

The sample estimate r_k is given below as :

$$r_k = \frac{\sum_{t=1}^{n-k} Z_t Z_{t+k} - \frac{1}{n-k}\left(\sum_{t=1}^{n-k} Z_t\right)\left(\sum_{t=k+1}^{n} Z_t\right)}{\left\{\sum_{t=1}^{n} Z_t^2 - \frac{1}{n-k}\left(\sum_{t=1}^{n-k} Z_t\right)^2\right\}^{0.5}\left\{\sum_{t=k+1}^{n} Z_t^2 - \frac{1}{n-k}\left(\sum_{t=k+1}^{n} Z_t\right)^2\right\}^{0.5}} \qquad \ldots.(5)$$

There are many different formulae for estimating unbiased r_k from the sample. The merits of different formulae for r_k are discussed in many research papers (Fiering and Jackson, 1971, Wallis, 1972, Jenkins and Watts, 1968, Kisiel, 1969). The estimate r_k given in Eq.(5) is by and large acceptable. Generally for low values of k, the differences are irrelevant and they are generally swamped by errors from the fundamental assumption of stationarity.

For an observed stationary time series, the values of r_k are unlikely to be zero, even if the corresponding values of ρ_k are zero. It is, therefore, essential to test the significance of r_k. Although an exact test of significance of r_k is not available, various approximate tests have been devised. One of the commonly used test is the one proposed by

Anderson (1942) which is based on the assumption that the time series is circular and the observations are normally distributed. The significance test used, for example, by Roesner and Yevjevich (1956) is based on this. For a serially independent series, E (r_k) are 0.0 and are approximately distributed as a normal variate with zero mean and variance as 1/n. Accordingly the test criterion is that if $|r_k| > 1.96/\sqrt{n}$ the hypothesis of independence is rejected.

There are also many direct tests like the turning points , phase lengths to determine linear dependence in time series which do not require the determination of r_k (Kendall, 1964, Matalas, 1967).

3.2.2 Spectral analysis

The correlogram analysis is in the time domain. A complementary method of finding the dependence in a time series is in the frequency domain by means of a spectrum, the term frequency being used in the harmonic sense and not in the histogram sense.

All stationary stochastic processes can be represented in the form

$$\mu_t = \int_{-\pi}^{\pi} e^{it\omega} dZ(\omega) \qquad \ldots(6)$$

Where $i = (-1)^{\frac{1}{2}}$ and $Z(\omega)$ is a complex random fluctuation.

It can be shown that the autocovariance ($\rho_k \sigma^2$) of a stationary process is

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\omega} dF(\omega) \qquad \ldots \quad (7)$$

Where k is the time lag, ω is the angular frequency and $F(\omega)/\gamma_o$ is a distribution fuction monotonically increasing and bounded between $F(-\pi) = 0$ and $F(\pi) = \sigma^2$. The function $F(\omega)$ is called the power spectral distribution function. For $k=0$, Eq. (7) is

$$\gamma_o = \sigma^2 = \int_{-\pi}^{\pi} dF(\omega) \quad \ldots (8)$$

In practical hydrologic application of spectral theory, Eq.(7) can be written as

$$\gamma_k = 2 \int_{0}^{\pi} \mathrm{Cos}(k\omega)\, f(\omega)\, d\omega$$

or

$$f(\omega) = \frac{1}{2\pi} (\gamma_o + \sum_{k=1}^{\infty} \gamma_k \mathrm{Cos}\, k\omega) \quad \ldots (9)$$

For a finite amount of data, an estimate of the power spectrum is

$$f'(\omega) = \frac{1}{2\pi} (C_o + 2 \sum_{k=1}^{T-1} C_k \mathrm{Cos}\, (k\omega)) \quad \ldots (10)$$

Where C_k is the autocovariance for a time lag k.

The estimate of the power spectrum by Eq. (10) is called the 'raw spectral estimate' because it does not give a smooth power spectral diagram. To adjust for smoothness, the smoothed estimate is commonly used in the form

$$F'(\omega) = \frac{1}{2\pi} \{ \lambda_o (\omega) C_o + 2 \sum_{k=1}^{m} \lambda_k(\omega) C_k \mathrm{Cos}\, k\omega \} \quad \ldots (11)$$

Where $\lambda_k(\omega)$ are selected weighting factor, m is a number chosen much smaller than T. A commonly used weighting factor is the Tukey-Hamming weights given as

$$\lambda_k(\omega) = 0.54 + 0.46\, \mathrm{Cos}\, (\pi k/m) \quad \ldots (12)$$

Where $m < \frac{T}{10}$

The significance of the spectrum is that it exhibits fewer sampling variation than the corresponding correlogram. Consequently, the estimated spectrum would provide a better evaluation of the various parameters involved in a model. If the process contains periodic terms, the frequencies of these terms will appear as high and sharp peaks in the estimated spectrum.

Extensive literature is available on the concept of spectrum and its statistical estimation procedure. (Lee, 1960, Bendat and Piersol, 1966). The use of spectral analysis to determine the non-random component in the time series have been done by many authors. For example, Julian (1967) used to determine the quasiperiodicities in precipitation, Landsberg et al (1959) investigation on water pollution in tidal estuaries. Wastler (1963) for the determination of dominant meander stream length, Adamowaski (1971), Carlson et al (1970), Andel and Balek (1971) for determining periodicities in hydrologic series. Fast Fourier transforms (FFT) have been used to estimate the predominant densities quickly. This approach becomes very handy when one is interested to determine hidden periodicities.

Between the correlogram analysis and the spectral analysis, which one gives a better view of internal dependence? For the diagnosis and estimation of linear stochastic models, which are commonly used in hydrological modelling, the correlogram has direct appeal and are easy to handle. However, for input-output processes and in spatially correlated systems ,spectral methods are more conducive to physical interpretation. Some comparative studies on the use of correlograms and spectral analysis are given by Chow and Kereliotis (1970) and by Kottegoda (1970).

The correlogram and the spectral analysis though used for identifying the existence of persistence in the series, yet, these are not the direct important hydrological parameters.

The two analysis which have direct bearing to the hydrological problems are (i) the run analysis and (ii) the range analysis. For example, if one has to determine whether droughts are persistent or not, the analysis of runs can be directly applied. Similarly, the storage capacity required to be estimated for a given set of inflows and demand can be directly studied from the range analysis. The computation of run lengths and rescaled adjusted range of a given series directly tells us whether the observed series is dependent or not. These properties for determining the dependence of the series is briefly described. An excellent review on these analysis is given by Yevjevich (1972).

3.2.3 Run analysis

In general a run is defined as a sequence of observations of the same kind preceded or succeeded by one or more observation of another kind. For example, the duration of time, the observed flows remains above or below a specified level (truncation level) is called the run length.

Corresponding to the runlength, the magnitude of the surplus or deficit is known as the run sum. Such properties of a time series are called as the crossing properties. The number of crossings at various truncation levels as well as the run length and run sum distribution are of special significanceto a hydrologist. These properties can also be used to check whether the series is dependent or not. If a median value is taken as the truncation level the mean run length, for an independent series, should be equal to 2. The series is dependent if the mean run length is greater than 2. A significance test for this also can easily be formulated.

An advantage of run analysis is that the dependence structure thus estimated is independent of the coefficient of skewness in the observed series. Correlogram analysis is valid strictly only for normal distribution (Sen, 1978).Another advantage is that differential persistence i.e. persistence at different truncation levels can be easily computed (Sen, 1978 , Chander et al, 1980).

3.2.4 Range analysis

For an observed sequence $Z_1, Z_2, \ldots\ldots Z_n$ with $E(Z) = \bar{Z}$, the range is defined by

$$R_n = \max_{1<i<n} \Sigma (Z_i - \bar{Z}) - \min_{1<i<n} \Sigma (Z_i - \bar{Z}) = d_n^+ - d_n^- \quad \ldots(13)$$

Range analysis is useful in the design of reservoir storage capacity. Take a hypothetical situation in which the observed flow sequence is routed through a reservoir of capacity R_n with initial storage d_n^- and a constant withdrawal rate of $\bar{Z}$. Under these conditions, it can be shown that the reservoir will be full on one or more occasions without overflowing, it will empty at least once and the withdrawal rate will be maintained throughout (Lawrence and Kottegoda, 1977).

In order to compare the results from different observed sequences, the range given in Eq. (13) is divided by the standard deviation of $Z_1, Z_2 \ldots\ldots Z_n$ to give the rescaled adjusted range.

$$R_n^* = R_n/\sigma_n \quad \ldots \quad (14)$$

A sequence can easily be checked to be dependent or independent as Hurst and Fuller (1951) showed that for an independent series R_n^*

$$R_n^* = (n/2)^{0.5} \qquad \ldots (15)$$

A value of the exponent larger than 0.5 in natural series means that the series is dependent. For independent series, range analysis has been studied extensively by Feller (1951), Anis and Lloya (1953, 75,76) Solari and Anis (1957) and Moran (1966).

3.3 Dependence Structure in Hydrologic Series

Given an observed series, dependence can be quantified by any of the methods given in the previous section. The most commonly used methods are the correlogram and the range analysis. However, lot of controversy has been generated on the structure of dependence when computed on the basis of these methods. For a stationary hydrologic series, the observed correlation coefficient r_k rapidly decreases with lag k. It implies that the events in the distant past have negligible influence on the present state of the process. This type of dependence is referred to as short memory dependence. Using range analysis most of the hydrological series has an exponent greater than 0.5 in Eq. (15) meaning thereby that the series is dependent. However, short term dependence cannot explain why the exponent is greater than 0.5? This can be explained only by long term dependence. Long term dependence is associated by the failure of the correlogram to die out meaning thereby that a process has infinite memory. There has been considerable controversy over long term dependence approach as pointed out by Mandelbrot (1969, 70), Chie et al (1973), Klemes (1974, 81), Mcleod and Hipel (1978). The controversy has been primarily generated based on the extensive studies on natural series by Hurst (1951). The short term dependence (derived from correlogram analysis), though having physical validity fail to satisfy the Hurst phenomenon. Whereas, long term dependence which does not seem physically realisable does statisfy the Hurst phenomenon. Thus there are two classes of models, long memory models and the short memory models. Long memory models are prescriptive type models ,whereas the short memory models are the descriptive type of models (Jackson, 1975). Before the details of these two classes of models (incorporating short and long memory dependence) are explained, the Hurst phenomenon is briefly explained.

3.4 Hurst Phenomenon

Hurst (1951, 57) did the range analysis on 690 annual time series, comprising streamflow, precipitation, temperature etc. It was shown that rescaled adjusted range (r.a.r) R_n^* varies with the length of record n as

$$R_n^* = (n/2)^k \qquad \ldots (16)$$

Hurst computed k for each of the time series over all the phenomenon studied, k was found to have an average value of 0.73 with a standard deviation of 0.08. The exponent k to have larger value than 0.5 (an independent series) is termed as Hurst phenomenon.

The four possible explanations for the Hurst phenomenon as given by Wallis and Matalas (1970) are:

i) Non normality of the probabilities distribution underlying the time series.

ii) Transience i.e. n not large enough for the Hurst coefficient to attain a limiting value of 0.5.

iii) Non stationarity in the observed series.

iv) Persistence in the time series.

For samples of moderate lengths, various simulation studies, (Matalas and Huzzen, 1967, Mandlebrot, 1969, O'Connel, 1977) have shown that R_n^* is very nearly independent of the distribution of the random variable. Hence the non-normality of the time series is to be the cause of the Hurst phenomenon can definitly be excluded.

With regard to hypothesis (ii), the implication is that if sufficiently long records are available in nature, k would tend to a value of 0.5 corresponding to asymptotic independence. Rejection or acceptance of this hypothesis must await the availability of longer geophysical records. At this stage (ii), and (iv) must be considered as either of them could be advanced as an explanation to the Hurst phenomenon.

Hurst (1957) suggested that a nonstationary model in which the mean of the series was subject to random changes could account for higher values of the Hurst coefficient k. Klemes (1974) has shown that even a zero order nonstationary model could reproduce the Hurst phenomenon. By simulation experiments with white noise, Klemes (1974), raised the mean level in different manner and showed how k increased due to this type of non-stationarity. But non-stationarity is rather an untractable assumption as it will be very difficult to fit non-stationarity properly to a given historical series.

It was therefore, conjectured that k having a value of 0.73 instead of 0.5 is due to persistence. The interest in the Hurst phenomenon increased manifold when it was observed that a value of $k > 0.5$ cannot be explained by Markov type of dependence identified through correlogram analysis. In fact all the short term dependent models belong to the Brownian domain wherein the estimate k tend to a value of 0.5 as the length of the sequence increases to infinity.

Hence Mandelbrot and others (1969) argued that if dependence is taken as the likely cause of Hurst phenomenon, then short memory models are inadequate and different class of models having infinite memory are needed.

Hurst phenomenon remains a puzzle as how does one explain physically that a geophysical process like riverflows has infinite memory? Infinite memory may be possible, say, in biological processes through genetic coding, than with processes related to inorganic nature like geophysical processes. In geophysical process, the memory seems to manifest itself mostly through conservation of mass and energy and has the Markovian property the past influence the future only through its effect on the present and thus once the present state has reached it matters little for the future development how it was arrived at.

4.0 STOCHASTIC MODELS

If by using correlogram analysis and the range analysis, it is observed that $\rho_k \neq 0$ and Hurst coefficient $H > 0.5$ within the significance range, the series is said to have persistence. If correlogram analysis is taken as the criterion of dependence a class of models which are fitted to the series is called as short memory models. Whereas, if the dependence structure is quantified through the Hurst coefficient, then the long memory models are used. These two classes of models are briefly explained.

4.1 Short Memory Models

4.1.1 Introduction

The use of short memory models in hydrologic analysis were introduced primarily to produce synthetic sequences of flows to route through a water resource system, the idea being to test it under a variety of conditions and with longer sequences of flows than historically available. The implication is that long sequences will contain more extensive events than observed and thus a more stringent test of the system. The basic requirement is that the synthetic flows should have properties which are indistinguishable from the historical flows, this is taken to mean that the statistical characteristics are maintained the same way as has been observed in the historical series.

A very early work on the use of short memory models is by Thomas and Fiering (1962) now famous as Harvard Water Programme (Mass et al, 1962). They introduced a monthly flow generator which is in effect a seasonal short memory model. It was applied in designing a water resources system for the Meramac river basin, Missouri, consisting of small reservoirs. Further application of such models include Fiering (1965), Hufschmidt and Fiering (1966), Schaake and Fiering (1967), Davis (1968), Hall et al (1969), Moreau and Pyatt (1970), Hamlin and Kottegoda (1971, 73), Gupta and Fordham (1972), Hamlin et al (1973,75), Spolia and Chander (1976) and many others. Most of these studies have not taken the sampling errors in the historical series into account. The theoretical implication of ignoring the sampling variability have not received the attention it deserves. Recently, these have been included through the Bayesian Framework analysis, notable studies being (Wallis and O'Connel, 1973, Lenton and Rodriguez Iturbe, 1974, Klemes, 1979).

Short memory models have also extensively been used for forecasting flows. If the series is non seasonal, the models used for simulation can also be used for forecasting. However, in the case of seasonal data, the special class of multiplicative time series model are preferred. The use of multiplicative models in riverflow forecasting is quite extensive and the notable studies being; Macmicheal and Hunter, (1972), Mekrecher and Delleur (1974, 1976), Clarke (1973) Delleur and Kavass (1978), Chander et al (1980). Recently, the use of Control Engg. concepts have been introduced in time series modelling. These are also called Bayesian forecasting (Harrison and Stevens, 1971, Maissis 1977, Chander et al, 1980). These models have also been used for extending the record (Hamlin and Kottegoda, 1971), infilling missing data (Kottegoda and Elgy, 1977), flood evaluation (Kottegoda, 1972, 73) etc.

4.1.2 Types of short memory models

The short memory models also called the ARMA (Autoregressive Moving Average) class of models is applied to the stochastic component, Z_t given in Eq (3). This means that the trend and seasonality are assumed to be removed and the residual is covariance stationary. In fact ARMA class of models is a special version of ARIMA (Autoregressive Integrated Moving Average) models which can directly account for the non-stationery behaviour of the series. ARIMA models can directly be applied to the series which has not been detrended or deseasonalised. An excellent review on ARIMA class of models is given by Box and Jenkins (1970).

AR(p) models: The p^{th} order linear autoregressive model suggest that the value Z_t at time t is constituted from the weighted sum of p values at times (t-1), (t-2) .... (t-p) and a random number, a_t

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t \qquad \ldots (17)$$

Where $E(Z_t) = 0$, $E(a_t) = 0$, $Var(Z_t) = \sigma_Z^2$ and

$Var(a_t) = \sigma_a^2$ and $E(a_t a_{t-k}) = 0$ and $E(Z_t a_{t-k}) = 0$.

MA(q) models: A moving average model of the order q is the one in which the current value of a random variable is the weighted sum of (q+1) random variables

$$Z_t = a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \qquad \ldots (18)$$

The stochastic component Z_t is equivalent in this case to the output from a linear filter with a white noise as input. In an infinite order moving average process, a necessary constraint for stationarity is that the sum of the squared weights should be less than infinity

Mandelbrot, 1976).

Autoregressive moving average, ARMA (p,q) model: The AR (p) and MA(q) models are special cases of ARMA (p,q) model given of the form

$$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j} \qquad \ldots (19)$$

The effect of linear aquifers and independent rainfall amounts justifies approximate representation of the river flow process through ARMA models. They are also analogous to those conceptual models in parametric hydrology that are based on linear reservoirs. For example, a nth degree linear reservoir system can be represented by ARMA (n, n-1) model. The relationship between parametric and stochastic hydrology using ARMA models is given by Moss (1972), Dooge (1972), Spolia and Chander (1974, 1979).

Box and Jenkins(1970) has introduced a backshift operator B in defining these models.

$$BZ_t = Z_{t-1}, \; B^p Z_t = Z_{t-p} \qquad \ldots (20)$$

Eq(19) can thus be written as

$$\phi_p^{(B)} Z_t = \theta_q^{(B)}(B)\, a_t \qquad \ldots (21)$$

where

$$\phi_p^{(B)} = (1 - \phi_1 B - \phi_2 B^2 \; \ldots\ldots\ldots - \phi_p B^p)$$

$$\theta_q^{(B)} = (1 - \theta_1 B - \theta_2 B^2 \; \ldots\ldots\ldots - \theta_q B^q)$$

A general class of ARIMA models is given in the next section.

4.2 Fitting of General ARIMA Class of Models

A multiplicate seasonal ARIMA model for Z_t series will be:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t^{(\lambda)} = \theta(B)\Theta(B^s)a_t \qquad \ldots (22)$$

where

s: seasonal length equal to 12 for monthly river flows

B: Backward shift operator defined by $B^sZ_t = Z_{t-s}$

a_t: Normally independently distributed white noise residual with mean equal to zero and variance σ_a^2

$\phi(B)$: $(1 - \phi_1 B - \phi_2 B^2 \ldots \phi_p B^p)$ non seasonal auto-regressive operator, where $\phi_1, \phi_2 \ldots \phi_p$ are non seasonal parameters

$\Phi(B^s)$: $(1 - \Phi_1 B^s - \Phi_2 B^{2s} \ldots \Phi_P B^{sP})$ seasonal AR operator of order P and Φ_i, i=1,2 .... P, are the seasonal AR parameters.

$\theta(B)$: $1-\theta_1 B-\theta_2 B^2 \ldots \theta_q B^q$; non seasonal moving average operator (MA), θ_i, i=1,2 .... q are the non seasonal MA parameters.

$\Theta(B^s)$: $1-\Theta_1 B^s - \Theta_2 B^{2s} \ldots \Theta_Q B^{sQ}$ are the seasonal MA operator of the order Q, Θ_i, i=1, ... Q, are the seasonal MA **parameters**.

$Z^{(\lambda)}$: Some appropriate transformation of Z_t such as Box Cox transformation (Box-Cox, 1964, Chander et al, 1980).

∇ : (1-B)

The notation (p,d,q) x (P,D,Q)s is used to represent the seasonal ARIMA model of Eq(22). The first set of brackets contains the order of the non-seasonal operators and the second pair of brackets has the order of the seasonal operators.

4.3 Steps used for Model Building

While applying a B-J model or in general any type of stochastic model to a particular problem it is recommended that the three stages of model development be adhered to. The first step is to identify the form of the model that may fit the given data. Once the model is tentatively identified the next step is to estimate the parameters of the model by an efficient parameter estimation technique. Once the parameters are estimated the model is checked for the possible inadequacies. If the diagnostic check reveals serious anomalies, appropriate model modification be made by repeating the identification and estimation stage. The ARIMA class of models building for a process is therefore a three stage iterative process consisting of:

1. Identification of the model
2. Parameter estimation
3. Diagnostic checking

4.3.1 Identification of the model

The purpose of the identification stage is to determine the order of differencing required to produce stationarity and also the order of both seasonal and nonseasonal AR and MA operators of the Z_t series.

The steps used for identifying the order of the model are:

i. Plot the original series: A visual inspection of a plotted time series may reveal one or more of the following characteristics i) seasonality ii) trends either in the mean level or in the variance of the series iii) persistence and iv) long term cycles.

ii. Autocorrelation function (ACF): To use the ACF in model identification, calculate and then plot r_k against lag k upto a maximum lag of roughly n/4.

Examine the plot of the ACF to detect the presence of non-stationarity in the Z_t series. When the data are non-stationary, differencing, is required. For seasonally correlated data with seasonal length equal to s, the ACF often follows a wave pattern with peaks, s, 2s, 3s and other integer multiple of s as shown in Box and Jenkins (1970, pp.174-175). If the ACF at lags that are integer multiples of the seasonal length s do not die out rapidly, this may indicate that seasonal differencing is needed to produce stationarity. Failure of ACF to damp out at other lags may imply that non-seasonal differencing is also required.

Once the data has been differenced enough to produce non-seasonal stationarity ($\nabla^d Z_t$) and both seasonal and nonseasonal stationarity ($\nabla^d \nabla_s^D Z_t$) of the seasonal data, check the ACF of the differenced series to determine the number of AR and MA parameters required in the model. Now the ACF of the differenced series is plotted.

If $\nabla^d \nabla_s^D Z_t$ is a white noise, the r_k is approximately NID (o, 1/n). Simply plot the confidence limit on the ACF diagram and check if significant number of r_k values falls outside the chosen confidence limit. When $\nabla^d \nabla_s^D Z_t$ is not a white noise then the following general rules may be invoked to help determine the type of the model required.

Non seasonal model: For a sure (o,d, q) process r_k cuts off and is not significantly different from zero after lag k. If r_k tails off and does not truncate this suggests that AR terms are needed to model the time series.

Seasonal model: When a process is a pure MA (o, d, q) x (0,D,Q)s model, r_k truncates and is not significantly different from zero after lag q + SQ.

PACF - Partial Autocorrelation function: The theoretical PACF can be calculated by using the Box -Jenkins approach (1970, Chapter 3). For model identification, plot the PACF coefficient ϕ_{kk} against lag k. The following general rule may prove helpful.

Non seasonal model: When the process is pure AR (p), ϕ_{kk} truncates and is not significantly different from zero after lag p. After lag p, ϕ_{kk} is approximately NID (o, 1/n).

Seasonal model: When the process is pure AR (p,d,o) x (P,D,0)s model , ϕ_{kk} cuts off and is not significantly different from zero after lag p + sP. After lag (p+sP), ϕ_{kk} is approximately NID (o, 1/n).

If ϕ_{kk} damps out at lags that are multiples of s, this suggest the incorporation of a seasonal MA component into the model.

Inverse autocorrelation function (IACF): Claveland (1972) defines the IACF of a time series as the ACF associated with the inverse of the spectral density function of the series. The IACF of the $\nabla^d \nabla_s^D Z_t$ series is defined by the ACF of (q,d,p) x (Q,D,P)s process. When the process is a pure AR process, r_k cuts off and is not significantly different from zero after lag p.

Inverse partial autocorrelation function (IPACF): IPACF is the inverse of the PACF and has the characteristics interchanged between the AR and MA process.

The ACF, PACF, IACF and IPACF transfer the given information into a format whereby it is possible to detect the number of AR and MA terms required in the model. In general, the ACF and IPACF truncates the pure MA process, while PACF and IACF cuts off the AR process.

For mixed process all four functions attenuate.

In order to get the details of the model identification the reader is advised to see the works of Hipel and Mcleod (1977), Ledolter (1978), Cline (1979), Chander et al (1980), Meckrecher and Delleur (1974) and Chatfield and Prothero (1973).

4.3.2 Estimation of parameters

If the parameters in the model are linear (AR models), they can be estimated by the use of Yule Walker equations or with the help of least square minimisation. However, these methods are not applicable when the parameters are non-linear.

In such cases, Box and Jenkins (1970, Chapter 7), suggest the use of approximate mle of the ARIMA model paramters be obtained by employing the unconditional sum of the squares method (Clarke, 1973). In this case, the unconditional sums of the squares function is minimised to get the least square parameter estimates. Recently Mcleod (1976) has described a modified sum of the squares method which provide the closest approximation to the Box-Jenkins exact maximum likelihood estimates.

When the moving average terms are present in the model, optimisation techniques are required to estimate the parameters. Some of the optimisation algorithms that have extensively been applied include the i) Gauss linearisation ii) the steepest descent and iii) the Marquardt algorithm. B-J have recommended the use of Marquardt algorthm (it has fast convergence even when the initial estimates are wrong) for the estimation of parameters. Chander et al (1980) have used the Marquardt algorithm in the estimation of ARIMA models parameters for the monthly flows of the Krishna and the Godavari river.

4.3.3 Selection of model

At the identification stage it is quite likely that not a single model is uniquely identified. In fact, two or three models are subjectively identified based on correlogram analysis and their parameters estimated. Now the problem of selecting the final model arises. Many times, model which gives the minimum variance in residuals is selected. This model selection rule can often lead to incorrect results. The main difficulty in the minimum variance rule is that it does not include the principle of parsimony of parameters.

One of earliest model selection rule including parsimony is based on the classical F-test in hypothesis testing (Astrom, 1967, Kashyap and Rao, 1976). Although this test weighs the order of the model in the decision, the test threshold is set by subjectively selecting an acceptable risk rate.

An approach not requiring arbitrarily specified parameters like significance levels has been proposed by Aikeke (1974). Based on information theoretic arguments, this information criterion is defined as twice the difference between the number of unknown model parameters and the maximum log likelihood. For ARIMA model the Aikeke Information Criterion (AIC) reduces to:

$$AIC = -2 \ln(\text{max. likelihood}) + 2k \qquad \ldots (23)$$

Where k is the number of AR and MA parameters used in the model. The model which gives the minimum AIC is finally chosen as the model. AIC criterion has been applied in hydrological time series by Mcleod and Hipel (1978), Cline (1979), Ozaki (1980). However, Mcleod studies showed that by using the AIC criterion, the model order increases as compared to minimum error variance model.

Another model selection criterion is the posterior probability (PP) criterion, developed independently by Schwarz (1977) and Kashyap (1977). The PP criterion also expresses parsimony but penalizes more heavily the extra parameters than the AIC criterion. It has been shown by Kashyap (1980) that PP criterion gives a more consistent decision rule for selecting a model than the AIC decision. Extensive literature exist in control Engg. journals about the time series model selection.

4.3.5 Diagnostic checking

Most diagnostic checks deals with the residuals assumptions in order to determine whether a_t are independent, homo scedastic and normally distributed. Residual estimates a_t are needed for the tests used in checking the three afore-mentioned residual assumption.
It may be mentioned that data transformation cannot correct dependence of the residuals because lack of independence indicated that the present model is inadequate. Rather the identification and estimation stages must be repeated in order to determine a suitable model.

Another class of diagnostic checking is done by over-fitting. Overfitting involves fitting a more elaborate model than the one estimated, to check if inclusion one or more parameters greatly improve the fit. For example, the PACF and the IACF may show decreasing but significant values at lag 1, lag 2 and at lag 9. If an ARMA (2,0) model is originally estimated, then a model to check by overfitting would be

$$(1-\phi_1 B - \phi_2 B^2 - \phi_9 B^9) Z_t = 0 \qquad \ldots (24)$$

A mle of ϕ_9, three or four times its standard estimate would definitely indicate that a more elaborate model should be selected. The likelihood ratio test (Mcleod, 1974) can also be used.

In order to determine whether the residuals are white noise an appropriate procedure would be to examine the residual auto-correlation coefficient. Another but less sensitive check is to calculate and to perform a significance test for the Portemantau Statistic.

5.0 COMMONLY USED SHORT MEMORY MODELS TO HYDROLOGIC TIME SERIES

5.1 Markov Autoregressive Model (AR 1)

Markov autoregressive model is perhaps the most commonly used model in stationary hydrologic time series. This model has frequently been applied to stationary annual series and to the standardised monthly series. In fact, most of the models fitted to hydrologic series, have not been derived, through the three stage iterative process of identification, estimation and diagnostic checking but have been derived heuristically, keeping in view the parsimony of parameters.

The first order Markov model was made popular by Fiering (1963). This model assumes that the entire influence of the past on the current value is reflected in the previous state value. The model has the form:

$$\frac{Z_t - \bar{Z}}{\sigma_Z} = \phi_1 \frac{Z_{t-1} - \bar{Z}}{\sigma_Z} + a_t \qquad \ldots (25)$$

It means that the departure of flow from the mean this year is a combination of a proportion given by ϕ_1 of the departure of the previous flows from its mean and a random component a_t. The a_t has zero mean and constant variance σ_a^2 and they do not depend on Z_{t-1} The distribution of a_t is same as that of the historical series Z_t. The two parameters, ϕ and σ_a^2 depend upon the statistical properties of the z_t series. If this model is required to model the mean, the standard deviation and the first order serial correlation ρ_1 then it can be easily shown that ϕ_1 should be equal to ρ_1 and $\sigma_a^2 = \sqrt{1-\rho_1^2}\, \sigma_Z^2$. Hence the AR(1) model of Eq. (25) can be written as

$$\frac{Z_t - \bar{Z}}{\sigma_Z} = \rho_1 \frac{Z_{t-1} - \bar{Z}}{\sigma_Z} + \sqrt{1 - \rho_1^2} \; a_t \qquad \ldots \quad (26)$$

where $E(a_t) = 0$ and $E(a_t^2) = 1$

The distribution of a_t affects the distribution of Z_t. If Z_t follows a normal distribution , a_t is then N (0,1) random deviates. First order Markov model in Eq.(26) has an inherent correlogram structure such that $|\rho_k| = |\rho_1|^k$

If Z_t is non normally distributed then a_t has to be non Gaussian. It is very important to model the non-Gaussian behaviour of the series because the Noah effect is directly dependent on the coefficient of skewness of the historical series. Fiering (1967) noted that this could be obtained by choosing a suitable skewness C_{sa} of the a_t series. For an annual series by cubing Eq.(26) and taking expectation the skewness C_{sa} of (a_t) is seen to be related to the skewness C_{sz} of the (Z) series by a relationship

$$C_{sa} = \left\{ \frac{1 - \rho_1^3}{(1 - \rho_1^2)^{3/2}} \right\} C_{sz} \qquad \ldots \quad (27)$$

Since $\left\{ \frac{1 - \rho_1^3}{(1 - \rho_1^2)^{3/2}} \right\}$ is always greater than or equal to one, it means that in order to maintain the coefficient of skewness, C_{sz} in the model, the coefficient of skewness C_{sa} of the random series is to be more than C_{sz}.

Knowing the value of C_{sa} , the next problem arises about the choice of the distribution of (a_t) having this C_{sa}. The precise choice of a family of distribution for (a_t) is more difficult and is often made by trial and error. In hydrologic time series analysis, the log normal and the gamma distribution have been frequently used. The gamma variable being generated by a χ^2 variable and by Wilson Hilferty (1931) transformation. If a_t is normally distributed with zero mean and unit variance, then a modified variate a'_t is defined by

$$a'_t = \frac{2}{C_{sa}} \left(1 + \frac{C_{sa} a_t}{6} - \frac{C_{sa}^2}{36} \right)^3 - \frac{2}{C_{sa}} \qquad \ldots(28)$$

and the first order Markov model becomes:

$$\frac{Z_t - \bar{Z}}{\sigma} = \rho_1 \frac{Z_t - \bar{Z}}{\sigma} + \sqrt{1-\rho_1^2} \; a'_t \qquad \ldots(29)$$

Eq.(29) now will model approximately the Z_t series which has mean $\bar{Z}$, σ , C_{sz} and ρ_1. However, McMahon and Miller (1971) showed this approximation was poor for skewness in excess of two which is often the case. Kirby (1972) suggested a modified transformation for Pearson type III distribution. Bernier (1970) and Weiss(1977) have obtained the exact distribution of (a_t) which ensures (Z_t) has gamma distribution, solution in other cases are awaited. It may be highlighted here that the choice of distribution of a_t directly affects the distribution of Z_t series. In fact, the distributions of (a_t) or (Z_t) uniquely specify each other.

5.2 Log Normal Model

In order to avoid distribution identification of the series from the samples, many a times, transformation of the data is done so that the transformed series is approximately normal. The stochastic model is then fitted on the transformed normal data. This aspect of transformation of data prior to modelling has been an important point of discussion (Beard, 1965, Pentland and Cuthbert, 1973). An important class of such transformation is the Box Cox transformation given as :

$$Z'_t = \frac{Z_t^{\lambda} - 1}{\lambda} \qquad \lambda \neq 0 \qquad \ldots(30)$$

$$= \ln Z_t \qquad \lambda = 0$$

There exist a unique value of λ ; which transforms the original series Z_t to Z'_t which is approximately normal. The procedure for estimating λ is given by Chander et al (1978). Generally hydrologists have been using the lognormal transformation which is a particular case of power transformation. The use of log normal transformation in hydrologic time series has been discussed in detail by Burges (1972) and Codner and McMahon (1973). The effect of such transformation in modelling is not well understood and so there had been considerable theoretical interest in the interrelationship between the parameters values and the behaviour of the model.

From the hydrologic modelling point of view it is important to note that if a transformation is applied to the data, it also changes the correlation structure. It is, therefore, important to determine the interrelationship between the correlation structure of the natural series to that of the transformed series. This relationship for a three parametric log normal distribution was first given by Yevjevich (1966) and the work has been further extended by Mejia and Rodriguez Iturbe (1974).

Let (Z_t) be the sequence of original flows which follows a three parametric log normal distribution i.e. (y_t) has a lognormal distribution with mean equal $\bar{y}$ and standard deviation σ_y. Now the mean, variance and skewness of (Z_t) can easily be related in terms of a, $\bar{y}$ and σ_y such as

$$y_t = \log(Z_t - a) \qquad \ldots(31)$$

then

$$\bar{Z} = a + \exp(\sigma_y^2/2 + \bar{y}) \qquad \ldots(32)$$

$$\sigma_z^2 = \exp \{ 2(\sigma_y^2 + \bar{y})\} - \exp (\sigma_y + 2\bar{y}) \qquad ...(33)$$

$$C_{sz} = \frac{\exp (3\sigma_y^2) - 3 \exp (\sigma_y^2) + 2}{\{ \exp (\sigma_y^2) - 1 \}^{3/2}} \qquad ...(34)$$

and

$$\rho_{1z} = \frac{\exp (\sigma_y^2 \rho_{1y}) - 1}{\exp (\sigma_y^2) -1} \qquad ...(35)$$

If the model is to maintain $\bar{Z}$, σ_z, C_{sz} and ρ_{1z}, Matalas (1966) suggested solution of four equations for calculating $\bar{y}$, σ_y, ρ_{1y} and a. Now a Markov model is fitted on the y- domain.

Mejia and Rodriguez (1974) showed that even by using these four equations only ρ_{1z} in the correlogram can be maintained. However, the complete correlogram structure, of the Z- series derived from the transformed Y series will be different from the one derived using directly the Z series. In the historical series the correlogram is decaying exponentially $\rho_{zk} = (\rho_{1z})^k$ whereas, the correlogram ρ_{yk} of the tranformed series will have a gradual decay of the form.

$$\rho_{yk} \simeq \frac{1}{\sigma_y^2} \exp (\sigma_y^2 - 1) \qquad ...(36)$$

It may be noted that Eq.36 has a asymptotic correlogram structure of the type of ARMA (1,1) process. In other words, an ARMA (1,1) process can be used to model (y) series and will result in a (Z) series which is log normal and have a Markovian structure.

5.3 Seasonal Markov Autoregressive Models

The Markov models discussed above assume that all time periods are characterised by identically distributed flows; the flows in each period have mean $\bar{Z}$, variance σ_z^2 and the lag one serial correlation ρ_{1z}. For studies that use only annual flows this assumption seems reasonable and is generally correct unless long term trends are present. However for studies that require flows for seasons, months or other sub-division of the year, more elaborate models are necessary. Certainly the mean flow for a month or other period during a wet season is different for the same duration in a dry season. Markov model can be modified to include the multiseason effect and a multiseason Markov model is used.

$$\frac{Z_{ij} - \bar{Z}_j}{\sigma_j} = \rho_j \frac{Z_{i,j-1} - \bar{Z}_{j-1}}{\sigma_{j-1}} + a_{ij}\sqrt{(1-\rho_j^2)} \qquad \ldots(37)$$

where i gives the number within the sequence of the year in which a given flow occurs and j indicates the season. In the case of monthly flows there will be 12 $\bar{Z}_j$ and σ_j respectively. In hydrology this type of first order single station model was originally used by Roesner and Yevjevich (1966). It is applied to monthly data by using the historical information to estimate $\bar{Z}_1$, $\bar{Z}_2$, ... $\bar{Z}_{12}$ and σ_1, σ_2, ...σ_{12} and choosing a distribution of a_{ij}. Application and discussion have been widespread and include Fiering (1964, 65, 67), Matlas (1967, b), Kardos (1975).

It is observed that even after standardising the series, the coefficient of skewness C_{sz} of each month is different. Similarly the degree of correlation ρ_j is not constant but goes on varying with season to season. A more general monthly model will therefore include all the 48 parameters (12 mean, standard deviation, skewness and lag one correlation coefficient). This is done by allowing ρ_j and a_{ij} distribution to depend on the position (month) of the season.

The generalised monthly flow model is written as:

$$\frac{Z_{i,j} - \bar{Z}_j}{\sigma_j} = \rho_{ij} \frac{Z_{i,j-1} - \bar{Z}_{j-1}}{\sigma_{j-1}} + \sqrt{1-\rho_j^2} \quad a'_{ij} \qquad \ldots(38)$$

where

$$a'_{ij} = \frac{2}{(C_{sa})_j} \left\{ 1 + \frac{(C_{sa})_j \; a_{ij}}{6} - \frac{(C_{sa})^2}{36} \right\}^3 - \frac{2}{(C_{sa})_j} \qquad \ldots(39)$$

and

$$(C_{sa})_j = \frac{(C_{sz})_j - \rho^3_{j-1} (C_{sz})_{j-1}}{(1 - \rho_j^2)^{3/2}} \qquad \ldots(40)$$

where $(C_{sz})_j$ is the coefficient of skewness in the observed series for the j^{th} month and a_{ij} is NID (0,1) and a'_{ij} is the modified random number with skewness correction.

Hall et al (1969) used this model in connection with critical periods of reservoirs. It was however, observed that estimation of monthly correlation and skewness from short records are unreliable (Hamlin and Kottegoda, 1971). Jones and Brelsford (1967) discusses autoregressive model with harmonically periodic parameters and employed then in a meteorological problem. However, incorporating month to month correlation and skewness of different months makes the model unwieldy and incorporates likely sampling errors in estimating these values from historical data.

Markov Models for Ephemeral Streams - One of the major problems in applying models to the ephemeral streams in arid lands is the modelling of zero flows. This is of particular importance in Indian context where some rivers experience zero flows during certain periods of the year. Fiering (1962) suggested a simple procedure of modelling zero flows. Taking a log normal AR(1) model for standardized monthly flow as:-

$$\frac{y_{t,j} - \bar{y}_j}{\sigma_j} = \rho_j \frac{y_{t,j-1} - \bar{y}_{j-1}}{\sigma_{j-1}} + \sqrt{1-\rho_j^2}\; a_{ij} \qquad \ldots(41)$$

$$y_{t,j} = \log (z_{t,j})$$

To model appropriate number of zero flows, the following procedure was adopted. For each month j the value of the normal deviate was obtained such that the area under the normal curve from $-\infty$ to k_j was equal to the proportion of p_j of zero flows in that month. Flows are modelled in the usual manner except that if the standard deviates in the log domain $(y_{t,j} - \bar{y})/\sigma_j$ were less than k_j the flows were set to zero. However, by using this procedure, it was found that the means and the standard deviations are underestimated. When three parameters log normal distribution was fitted to the data and the same procedure followed, then it results in an overestimation of mean and producing large number of zero flows. Jackson (1975) suggested the use of Birth-death Markov chain models . Srikanthan and McMohan (1980) suggested the method of fragments for modelling monthly flows for ephemeral streams. However, much work still remains to be done in this direction.

5.4 Higher Order Autoregressive Models

It is quite likely that the Markov models do not always model the streamflow series satisfactory. Higher order autoregressive models will be required in such cases where the correlogram structure in the observed series does not decay in the exponential form. This may become necessary to accommodate runoff conditions in which the groundwater aquifer stores water from season to season and then contributes a fraction to each season from the total runoff. The behaviour of ground water storage may thus necessitate the use of higher order models. In the case of AR(1) model, it is assumed that $\rho_{|k|} = \rho_1^{|k|}$. However, in the observed series the correlogram structure may be entirely different from $\rho_{|k|}$. The inclusion of the observed correlogram structure at greater than one lag may require the use of higher order models. Higher order (AR) models have been applied when the Markov model was found inadequate and when there was no evidence of low frequency effects in the series. The AR(p) model can be written as:-

$$\frac{Z_t - \bar{Z}}{\sigma_z} = \phi_1 \frac{Z_{t-1} - \bar{Z}}{\sigma_z} + \ldots\ldots\ldots \phi_p \frac{Z_{t-p} - \bar{Z}}{\sigma_z} + \sqrt{1-R^2}\, a_t \quad \ldots(42).$$

R^2 : Degree of determination given as

$$R^2 = (\phi_1\rho_1 + \phi_2\rho_2 + \ldots\ldots\ldots \phi_p\rho_p)$$

A properly evaluated AR(p) model will maintain mean, standard deviation and ρ_1 to ρ_p correlatin coefficients. The skewness can easily be incorporated in a similar way as done in AR(1) model. The identification, estimation of parameters and the diagnostic checking procedure remaining the same as given in section (4.3). Fiering (1967) used a 20 lag model in order to satisfy the Hurst phenomenon. Generally, higher order models have been used for daily and pentad flows (5 daily) series. For example, Quimpo (1967, 68) fitted a second order model on standardized daily flows to 17 American rivers; the standardization was achieved by fitting harmonic function to 365 daily means and standard deviations. This model was validated on the basis of correlogram and spectrum analysis on both the historical and residual series and on the ratio of the variances of the standardised and residual series, but no study was done on the skewness coefficients and about the distributions of the series. The pentad flow data of five English rivers was analysed by Kottegoda (1972) and the model fitted was fourth order autoregressive model. It was however observed that large coefficient of skewness exist in the pentad flow series which required the use of Pearson III or VI distribution. Spolia and Chander (1977) used Cannonical expansion on the monthly flows of the Sutlej and the Beas river. The advantage of Cannonical model is that it tries to maintain the whole correlation matrix between months. However their model used the log transformation which could not maintain the coefficient of skewness of each month. Phien (1979) suggested a procedure which can maintain the coefficient of skewness of each month. This required the use of gamma variable and a simple modification of scheme provided by Whitakar (1973). Tao and Delleur (1976) used a time varying non seasonal ARMA model for modelling monthly flows, where the skewness was accounted by using the logarithmic transformation and the parameter ϕ_j and θ_j of an ARMA model were dependent on the ρ_j values. Tao and Delleur (1976) also suggested a modified Portmanteau Q statistic for checking the seasonal correlation coefficient for residual whiteness .

All these higher order models were used on the standardised monthly or pentad series. However, multiplicative models can be applied directly to the series without being standardised. Chander et al (1980) modelled the monthly flow series of the Krishna river and the Godavari river by multiplicative ARIMA models. The models identified were

$(1,0,0) \times (0,1,1)_{12}$ model for the Krishna river and $(2,0,0) \times (0,1,1)_{12}$ model for the Godavari river. The generalised Box-Cox transformation was applied and the parameters of the model and the Box-Cox transformation λ were determined by modified Marquardt algorithm. Similar, multiplicate models were also used by Clarke (1973), Mekrecher,etal(1976)& Delleur and Kavaas (1978). This model is parsimonious and does not disaggregate the series into seasonal and nonseasonal components. Other studies involving higher order models are by Beard (1967 b), Adamowaski (1971). However, further studies on the relationship between the distribution of the standardised (Z_t) series with the distribution of the (a_t) series needs to be done.

5.5 Daily Flow Studies - Recent Studies

The modelling of realistic daily flow sequences is slightly complex because of three reaons, viz., (i) Daily flows are highly non Gaussian, large variability and having high ρ_1 correlation between flows, (ii) there are spurts of rising limbs followed by a longer period of falling flows (iii) there are many days when the flow is exactly zero. This asymmetrical behaviour of hydrographs leads to the statistical property of time irreversibility. The autoregressive models described earlier are based on the principle of time reversibility and therefore can't produce sharp rising limbs and slow recession. A general belief is that monthly flow models which frequently employ a logarithmic transformation, so that a linear Gaussian process can be applied to model the transformed process, can be readily applied to model daily flows is generally incorrect. Recession effects which are frequently evident in daily flows but which tend to be averaged out in monthly and annual flows cannot be reproduced by any linear model based on a Gaussian distribution (Weiss, 1973).

The early daily flow modelling of Beard (1976 b) and Quimpo (1967, 68) did not reproduce this rising - falling behaviour. Beard first generates the monthly flows and then disaggregates them during a flood season using a second order Markov Chain and the frequency characteristics of daily flows within a calender month. Daily flows are then adjusted to agree with simulated monthly flows. Payne et al (1969) extended this work for a multisite modelling. A more general disaggregating model is described by Valencia and Schaake (1973) and by Mejia and Rouselle (1976). All these approaches tries to rearrange the flow series generated by other means empirically so that the number of recession during a particular month are reproduced. However, this approach is quite adhoc and does not produce ascension-recession behaviour satisfactorily. A daily flow model based on non-linear type autoregressive model is also given by Yakowitz (1973). This model is operationally successful in modelling zero flows and steep rising and falling behaviour. However, this model lacks mathematical tractability. In this connection, Weiss (1977) suggested a new model known as 'Shot Noise Model' which has a built in capacity to model the ascension recession behaviour. This model has great potentialities and is briefly explained.

5.5.1 Daily flow shot noise model

When modelling a stochastic process in hydrology a linear assumption is generally made

$$Z(t) = \int_0^\infty h(u)\, dY(t-u) \qquad \ldots(43)$$

Where $dY(t)$ is a process of independent and hence uncorrelated increments which describe all the randomness in $Z(t)$, and $h(t)$ is the system transfer function. Obviously, $Z(t)$ is the flow at time t which comes from the rainfall $Y(t)$. The choice of the transfer function $h(t)$ determines the autocorrelation function of $Z(t)$. The particular $h(t) = e^{-bt}$ gives a AR(1) process and corresponds to a single linear reservoir.

Weiss (1973,77) showed that the recession shape would be produced in the daily flows if $dY(t)$ (rainfall) is taken to be zero almost every where except for a delta function at random times. This is possible if Y is the value of the shot (rainfall) which is a random variable with an exponential distribution of mean θ, independent from shot to shot. Since the rainfall occurs only for few days in a year, its occurrence is taken to be Poisson distribution at a rate of ν. Each rainfall event contributes to run off $Z(t)$ as Ye^{-bx} where x gives the arrival time of the rainfall shot prior to t and b is the decay rate.

If $N(t)$ is the number of rainy days (shots) and τ_m denotes the arrival time of the shots, the river flow $Z(t)$ is

$$Z(t) = \sum_{m=-\infty}^{N(t)} Y_m \exp\{-b(t-\tau_m)\} \qquad \ldots(44)$$

where Y is an exponential distribution having mean θ, of p.d.f.

$$\frac{1}{\theta} \exp(-y/\theta) \quad \text{for} \quad y > 0.$$

The process has three parameters, the rainfall event rate ν, the magnitude of rainfall θ and the decay rate of run off $Z(t)$ if there is no rainfall is b.

It can be shown that $Z(t)$ has a gamma distribution with parameters

$(1/_\theta , \nu /b)$ and thus Z(t) is non-negative and positively skewed. The statistical parameters of Z(t) are

$$E(Z(t)) = \nu\theta /b \qquad ...(45)$$

$$Var\ (Z(t)) = \nu\theta^2/b \qquad ...(46)$$

$$\rho_k(Z(t)\ Z(t+k)) = e^{-bk} \qquad ...(47)$$

Equation 44 can be writtern as

$$Z(t+k) = e^{-bk} \sum_{m=-\infty}^{N(t)} Y_m \exp\ (-b(t-\tau_m))$$

$$+ \sum_{m=N(t)}^{N(t+k)} Y_m \exp\ (-b\ (t+k-\tau_m)) \qquad ...(48)$$

The two terms in equation (48) are independent. The first represents the effect of rainfall events previous to t and is equal to $e^{-bk}Z(t)$ and the second includes rainfall during time t and (t+k) and is the innovation term. Denoting the innovation by a(t+k)

$$Z(t+k) = e^{-bk}Z(t)+a(t+k) \qquad ...(49)$$

The shot noise model is thus the first order AR process in continuous time. It differs from AR process that a(t+k) is non-negative (instead of being Gaussian) with a skewed distribution and with a positive probability of being exactly zero. Weiss (1977) also suggested a double shot noise model, one to represent direct runoff for rainfall and to account for groundwater storage. If used in seasonal flow modelling, obviously, the value of ν , θ , b may vary from month to month. This may need estimation of 36 parameters from the observed data. Further development in relating the model and its physical characteristics are awaited as also the seasonal version with an economical set of parameters.

6.0 MULTISITE SHORT MEMORY MODELS

All the models considered so far deal with modelling of flows at a single site. Multisite models are required if there are many tributaries in the catchment or when the catchment to be modelled is very large having many river gauging sites. In such cases it is not sufficient simply to model flows at every site independently because the flows at various sites can be strongly interrelated. For example, if the flow is high at one time in a particular stream then it will tend to be high sometime later in a lower reach of the stream. Independent modelling of river flows for multiple sites can not preserve spatial and temporal correlations between flows, consequently, multivariate modelling which accounts for serial and cross correlations are needed (Matalas, 1971, Fiering 1964).

6.1 Multisite Lag One Autoregressive Model

Assume there are m sites for which stochastic river flow model is to be formulated. The characteristics to be modelled are the mean, the standard deviation at each site as well as lag one serial correlation and lag zero cross correlation between each site.

Let Z_{ij} be the flow during the i^{th} period at the j^{th} station and $\bar{Z}_j$, σ_j, $(C_{sZ})_j$ and $\rho_{i,j}$ be the mean, standard deviation, skewness and the lag one serial correlation for the flows at j^{th} station. Also let ρ_{0jk} be the zero cross correlation between j and k^{th} station. Then the flows at different stations (1,2....... m) at time (t+1) will be related to the flows at time t by a matrix equation

$$
\begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_{t+1} = \begin{bmatrix} \\ \\ A \\ \\ \\ \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_t + \begin{bmatrix} \\ \\ B \\ \\ \\ \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix}_{t+1} \quad \ldots(50)
$$

Where A and B are (m x m) matrices of parameters and a_{t+1} is a (mxl) vector of independent stochastic components which are also independent of Z_t. If Z_i series is standarised the matrices A and B can be written in terms of lag one and lag zero cross correlation coefficients. This can be shown as follows:

$$
\text{Let} \quad \left[M_o\right] = E\left[\left[Z\right]_t \left[Z\right]_t^T\right]
$$

$$
\left[M_o\right] = E \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_t \left[(Z_1, Z_2 \ldots\ldots\ldots\ldots Z_m)\right]_t
$$

$$
\left[M_o\right] = \begin{bmatrix} 1 & \rho_{0,12} & \cdots\cdots & \rho_{0,1m} \\ \rho_{0,21} & 1 & & \rho_{0,2m} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \rho_{0,m1} & & & 1 \end{bmatrix} \quad \ldots(51)
$$

where $\rho_{0,ij}$ is the lag zero cross correlation coefficient for station i and j.

$$
\text{Similarly let} \quad \left[M_1\right] = E\left[\left[Z\right]_{t+1} \left[Z\right]_t^T\right]
$$

$$\left[M_1\right] = \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_{t+1} \left[(Z_1, Z_2 \ldots\ldots\ldots\ldots Z_m)\right]_t$$

$$= \begin{bmatrix} \rho_{11} & \rho_{12} & & \rho_{1m} \\ \rho_{21} & \rho_{22} & & \rho_{2m} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ \rho_{m1} & & & \rho_{mm} \end{bmatrix} \qquad \ldots(52)$$

where ρ_{ij} is the lag one serial correlation between station i and j.

The matrix $\left[M_o\right]$ given in equation (51) represents the lag zero cross correlation between m-sites and $\left[M_1\right]$ represents the lag one cross correlations of the stations with diagonals having lag one serial correlations.

Equation (50) can be written as:

$$\left[Z\right]_{t+1} = \left[A\right]\left[Z\right]_t + \left[B\right]\left[a\right]_{t+1} \qquad \ldots(53)$$

or

$$E\left[\left[Z\right]_{t+1}\left[Z\right]_t^T\right] = \left[A\right]\left[Z\right]_t\left[Z\right]_t^T + \left[B\right]\left[a\right]_{t+1}\left[Z\right]_t^T$$

$$\left[M_1\right] = \left[A\right]\left[M_o\right]$$

$$\left[A\right] = \left[M_1\right]\left[M_o\right]^{-1} \qquad \ldots(54)$$

For which $\left[A\right]$ can be estimated by substituting the estimated correlation coefficients in the matrices $\left[M_1\right]$ and $\left[M_o\right]$. It can also be shown that

$$\left[B\right]\left[B\right]^T = \left[M_o\right] - \left[A\right]\left[M_1\right]^T \qquad \ldots(55)$$

Unlike $\left[A\right]$ which has a unique solution, $\left[B\right]$ does not have a unique set of solution. An infinite number of $\left[B\right]$ exist that satisfy equation (55) such that BB^T is positive definite matrix (Slack, 1973; Young, 1968). Matalas (1967) following Fiering (1964) suggested the use of method of equivalence and the principal component analysis. The unique solution of $\left[B\right]$ is possible only if the structure of the matrix $\left[B\right]$ is triangular. Restricting the $\left[B\right]$ form as triangular will restrict the correlation constraints in the process which has not been explored. Slack (1973) has further shown that sometimes it may be impossible to substitute the historically computed serial correlation and cross correlation values in the $\left[M_o\right]$ and $\left[M_1\right]$ matrices, as by doing so it may violate the positive definiteness properties of $\left[B\right]$ $\left[a\right]_t$ implying that the determinate of $\left[B\right]\left[B\right]^T < 0$ (Matalas and Wallis, 1971 b). Maximum likelihood estimators of parameters in a two site model from unequal length of historical records have been attempted by Frost et al ('73). Fiering ('68) & Crosby et al ('70) suggested modified mean of calculating M_o & M_1. This may still lead to unacceptable parameter sets. The preservation of skewness and the use of multi lag models have been done by Matalas and Wallis (1974) and Ledolter (1978).

As an application of multisite model, let us take the example of two station flows. Taking the triangular form of the matrix B, the model given in equation (50) can be written for the standardized Z series as:

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}_{t+1} = \begin{bmatrix} \rho_{11} & 0 \\ 0 & \rho_{12} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}_t + \begin{bmatrix} \alpha^2 & 0 \\ \frac{\rho_{0,12}\ \beta}{\alpha} & \left(\frac{\alpha^2\ \gamma^2 - \rho_{0,12}\ \beta^2}{\alpha^2}\right)^{0.5} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}_{t+1} \quad \ldots(56)$$

where,

$$\alpha^2 = 1-\rho^2_{11}, \quad \beta = 1 - \rho_{11}\ \rho_{12}; \quad \gamma^2 = 1 - \rho^2_{12}$$

Rodriguez-Iturbe (1969) has fitted the two site model given in equation (56) to two river flow sequences. Many a times two station models have been used where one station is the key station and the second station is the satellite station. This key and satellite station model was first suggested by Fiering (1964) which were applied to the Great lake by Megerian and Pentland (1968). In such a case it is assumed that the flows at the key station follows a Markov process whereas the flows at the satellite station is a combination of certain proportion of flows at the previous time at satellite station, the flow at the previous time at the key station and an independent stochastic component. Lawrence (1976) improved upon the Fiering model in determining the matrices $[A]$ and $[B]$ such that the lag 0 and lag 1 cross correlation coefficients are maintained in the model.

6.2 Higher Order Short Memory Multisite Models

Higher order multisite models will be required if serial and cross correlation coefficients at lags more than one between sites is to be maintained. Thus higher lag models may be required in the case of weekly and pentad flows where the correlation structure will not be markovian. A higher lag multisite model will have the form:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_{t+1} = \begin{bmatrix} \\ A_1 \\ \\ \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_t + \cdots\cdots\cdots \begin{bmatrix} \\ A_p \\ \\ \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_m \end{bmatrix}_t$$

$$+ \begin{bmatrix} \\ B \\ \\ \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix}_t \qquad \ldots(57)$$

Obviously the matrices $\begin{bmatrix} A_1 \end{bmatrix} \ldots\ldots \begin{bmatrix} A_p \end{bmatrix}$ will depend upon the serial correlation and cross correlations at various lags. The daily flow model of Pentland and Cuthbert (1973) is a modification of Beard (1965) multisite monthly flow model and can be considered to be a special case of equation (57). However because of large number of coefficients to be estimated, the results will be very much data dependent, unstable and furthermore they will be of questionable reliability.

6.3 Decoupled Multisite Models

The classical multisite modelling as described in the previous section is quite cumbersome because of simultaneous estimation of parameters in the $[A]$ and $[B]$ matrix. In order to simplify the estimation procedure, decoupled multisite models have been evolved by Ramaseshan (1975). In this methodology, the modelling is done at two stages. First, a suitably identified ARMA (p,q) model is fitted to the standardised and the transformed (for accounting skewness) series at each site. The fitting is done by treating the process as univariate series. After the univariate model at each site is fitted, the serially independent random component at each site is separated and tested for their randomness. This is done to ensure that a properly identified and validated ARMA model has been fitted to the data. Now the residual series a_{tj}, will be serially independent but will have cross correlation with the residual series a_{tk} at other stations. Now, again an ARMA (p,q) model is fitted to the cross correlated residual series such that after the model fitting, the residuals are white noise. By coupling the two models, a coupled multivariate model is obtained.

For example, if it is assumed that the standardised series at various sites follows an AR(p) process, then a multivariate model can be decoupled as follows:

First a univariate series model is identified:

$$[Z]_{t+1} = [A_1][Z]_t + [A_2][Z]_{t-1} \cdots + [A_p][Z]_{t-p} + [\eta]_{t+1}$$

Where $[A_i]$ are the diagonal matrices of autoregressive components. After the univariate model is fitted, the random quantities $[n]_{t+1}$ which will be cross correlated are modelled to satisfy the cross correlation matrix between residuals:

$$[n]_{t+1} = [C][n]_t + [D][a]_{t+1}$$

The model is adequate if $[a]_t$ is purely random and free from serial and cross correlation. Ramaseshan (1975) has used the decoupled multivariate model for maintaining the serial and cross correlation matrix in daily flows at three tributaries sites with good result. Decoupled multivariate modelling may prove extremely efficient in multisite analysis. However, further work is required in the unbiased estimation of $[C]$ and $[D]$ matrices.

7.0 LONG MEMORY MODELS

The importance and genesis of long memory models have been discussed in section (3.3). In this section, for and against arguments for the use of long memory models are not discussed, the survey by Klemes (1974) is quite exhaustive. What is attempted in this section is how the long memory effect is incorporated in a stationary process. The most developed type of such models are the fractional noise models of Mandelbrot and Wallis (1968, 1969). Also there are Broken line models of Mejia et al (1972, 1974). Recently Weiss (1977) provided a fast fractional version of shot noise models. These models are briefly explained in the following sections.

7.1 Fractional Gaussian Models

All the short memory models follow a Brownian behaviour. A random motion B(t) is said to be Brownian if:

(i) $B(t + u) - B(t)$ is $N(Q,u)$ ...(58)

(ii) $B(t+u) - B(t)$ and $\dfrac{B(t+ur) - B(t)}{r^{0.5}}$ have the same distribution.

A Brownian motion satisfies the asymptotic $n^{0.5}$ law as given by Hurst (1957).

Since in a long memory model, the value of Hurst h is to be greater than 0.5, it is assumed that the fractional Brownian motion is composed of moving average of previous increments weighted by $(t-u)^{h-0.5}$ where h is between 0.5 and 1.0. Also the FBM is self similar but it follows

r^h law where $h > 0.5$ i.e.

(i) $$B_h(t) = \int_{-\infty}^{t} (t-u)^{h-0.5} \, dB(u) \quad \ldots \; 0.5 < h < 1.0 \qquad \ldots(59)$$

(ii) $B(t+u)-B(t)$ and $\dfrac{B(t+ur)-B(t)}{r^h}$ have the same distribution.(FGN model has been criticised on this self similarity property by Scheideggerr 1970). Obviously the condition suggests a long memory model (suitably weighted) if $h > 0.5$. If $h = 0.5$ fractional Brownian motion follows a Brownian motion.

A discrete fractional Brownian motion (dt FBM) is a discrete derivative of FBM and is given as

$$b_t(h) = B_h(t) - B_h(t-1) \qquad \ldots(60)$$

Now a river flow $b_t(h)$ is said to follow discrete fractional Brownian motion having a correlation structure defined by Cor $\rho(b_t(h)\, b_t(h+k))$.

By applying the principle of self similarity, it can be shown that the correlation structure of dt FBM is given as (Lawrence and Kottegoda, 1977).

$$\rho\{b_t(h)\, b_t(h+k)\} = \frac{1}{2}\{ (k+1)^{2h} - 2k^{2h} + (k-1)^{2h}\} \qquad \ldots(61)$$

when k is very large; Equation (61) is simplified as:

$$\rho\{b_t(h)\, b_t(h+k)\} = \{h(2h-1)\}\, k^{2h-2} \quad \text{for } k >> 0 \qquad \ldots(62)$$

As can be observed from equation (61) and (62) the serial correlations of dt FBM are positive for all lags when $1/2 < h < 1.0$, they will sum upto infinity as indicative of the long memory. If h=0.5 the process is random. It can be observed from equation (61) and equation (62) that the correlogram does not decay exponentially but vary gradually.

7.2 Modelling of FGN Process

Historically, researchers did not develop an exact modelling of FGN process, because the computation of a single FGN variate involves an infinite number of operations. The methods used for obtaining approximate realisation of FGN include the Type I and Type II approximation of Mandelbrot and Wallis (1969 C) which are finite moving averages in which the number of Gaussian variables that are averaged is proportional to the size n of the desired sample. The type I process which follows the theoretical form of the FGN process is expensive in computer time and is cumbersome to compute . The type II approximation is deficient in high frequencies and its low frequencies approximation is satisfactory only when h is close to one. Matalas and Wallis (1971) proposed a filtered FGN model which improves the high frequency behaviour of the type II approximation but it still takes a large amount of computer time. The fast fractional Gaussian (FFGN) generator proposed by Mandelbrot (1971) is by the most efficient approximation and is briefly described here.

7.2.1 Fast fractional Gaussian noise model

The FFGN variates $(Z_f(t,h))$ are obtained by summing about a long memory Markov process $Z_l(t,h)$ and a short memory one $Z_h(t,h)$ as follows:

$$Z_f(t,h) = Z_1(t,h) + Z_h(t,h) \qquad \ldots(63)$$

The low frequency term is defined as

$$Z\ (t,h) = \sum_{m=1}^{L} W_m \ Z(t,\rho_m) \qquad \ldots(64)$$

in which $Z(t,\rho_m)$ is a first order Markov process with zero mean and unit variance having lag one autocorrelation coefficient ρ_m, $\rho_m = \exp(-B^{-m})$. For example, when $m = 1$, $Z(t,\rho_m)$ is an AR(1) process with $\rho_1 = 1$ and when $m = 10$, $Z(t, \rho_{10})$ is an AR(1) process with $\rho_1 = \exp(-B^{-10})$.

W_m is the weightage **factor which** is given as:

$$W_m^2 = \{h\ (2h-1)(B^{1-h} - B^{h-1}) \ / \ \Gamma(3-2h)\} \ B^{-2(1-h)m} \qquad \ldots(65)$$

The parameter B can take a value between 2 to 4 and L ranges from 15 to 20. Together B and L determines the quality of the approximation (Mandlebrot, 1971).

The high frequency term is the first order Markov process of mean equal to zero, but a variance σ_h^2 or $(1 - \sum_{m=1}^{L} W_m^2)$ having a lag one correlation coefficient of

$$\rho_{1h} = \frac{\rho_1 - \sum_{m=1}^{L} W_m^2 \ \rho_m}{(1 - \sum_{m=1}^{L} W_m^2)} \qquad \text{where} \qquad \rho_m = \exp(-B^{-m}) \qquad \ldots(66)$$

in which ρ_1 is the lag one autocorrelation coefficient required in the stationary process.

7.3 FFGN Model Accounting

Modelling of skewness in the FFGN process can be done in two ways:

i) Modify only high frequency term : Normally distributed random numbers are used for the low frequency term while skewed random numbers are used for the high frequency term. The skewness of the random number is related to the observed skewness in the series by the **Wilson-Hilferty** transformation:

$$C_{sa} = \frac{1 - \rho_h^3}{\sigma_h^3 \, (1 - \rho_h^2)^{3/2}} \, (C_{sz}) \qquad \ldots(67)$$

ii) Modify both high and low frequency term: In this case skewed random number is used for both high and low frequency factors and their skewness is given by

$$C_{sa} = C_{sz} \left[\frac{\sum_{m=1}^{L} W_m^3 \, (1 - \rho_m^3)^{3/2}}{\dfrac{1 - \rho_m^3 + \sigma_h^2 \, (1 - \rho_h^2)^{3/2}}{(1 - \rho_h^3)}} \right]^{-1} \qquad \ldots(68)$$

where $\rho_m = \exp(-B^{-m})$

iii) Another alternative to generate skewed flows is to use the lognormal transformation. Burges and Lattenmaier (1977) investigated the distortion resulting in the autocorrelation function of a FFGN process and provided a set of tables for choosing an appropriate value of the Hurst coefficient h in the log domain so that the skewness in the model is properly implemented.

7.4 Exact Modelling of FGN Process

Hipel and Mcleod (1978) has suggested an improved methodology for the exact modelling of a FGN process. A FGN model can be specified by three parameters, viz, $\bar{Z}$, σ_Z and Hurst h. The theoretical autocorrelation coefficient is given by

$$\rho_k = \frac{1}{2} \{ (k+1)^{2h} - 2k^{2h} + (k-1)^{2h} \} \qquad \ldots(69)$$

The procedure for exact FGN modelling is as follows:

Suppose a series $Z_1, Z_2 \ldots\ldots\ldots Z_n$ with parameters $\bar{Z}$, σ_Z and Hurst h is to be modelled.

i) Model a Gaussian random sequence

$(a_1, a_2 \ldots\ldots\ldots\ldots a_n)$ having NID (0,1)

ii) Calculate the (n x n) correlation matrix

$$\left[C_n(h)\right] = \left[\rho_{|i-j|}\right] \qquad \ldots(70)$$

$\rho_{|i-j|}$ computed from equation (69) and $\rho_o = 1.0$

iii) Use the Choleksy decomposition (Hearly 1969) of $\left[C_n(h)\right]$ such that

$$\left[C_n(h)\right] = \left[M\right]\left[M\right]^T \qquad \ldots(71)$$

where $\left[M\right] = \left[m_{ij}\right]$ n x n lower triangular matrix.

The exact simulation of FGN is then simply

$$\frac{Z_t - \bar{Z}}{\sigma} = \sum_{i=1}^{t} m_{ti} \, a_i \qquad \ldots(72)$$

Now the Z_t modelled in equation (72) will be a FGN series as $N(\bar{Z}, \sigma_Z)$. Obviously if a series of n values is to be modelled, it requires a computer storage capacity of $\frac{n}{2}$ (n+1) corresponding to $\left[C_n(h)\right]$. In case skewed FGN process is to be modelled then the random variate a_t has to be skewed which can be achieved using Wilson Hilferty transformation.

7.5 Estimation of FGN Parameters

The modelling of FGN process depends upon three parameters, i.e mean, standard deviation, and the Hurst h. The value of the Hurst h, then determines the correlation structure of the process. The exact determination of Hurst h from the historical series is therefore very important. There exist a lot of ad hocism in the estimation if h and a better method of realistic estimation of h is required (Lawrence and Kottegoda, 1977). The commonly used method for determining h is by the parameter k given in equation (16). Matalas and Huzzen (1967) presented a table for mean values of k for various values of n and ρ for a Markovian process.

Another estimation of h is h* suggested by Mandelbrot and Wallis (1969 c). In this method, a time series of n observations is divided into a subset of time series, referred to as subseries of length n_1, where $S < n_1 < n$. For each subseries of length n_1, the standard deviation S_{n1}, and the rescaled adjusted range R_{n1}, are determined. The slope of R_{n1}/S_{n1} versus lag n_1 is denoted as h* and is the estimate of h.

However, both h* and k estimators have rather poor properties, particularly in the case of annual time series of river flows which have small size (O'Connell, 1977).

Mcleod and Hipel (1978) estimated the maximum likelihood estimate of h based on the study of Dansumir and Hanan (1976) as follows:

Given a historical time series $Z_1, Z_2 \ldots\ldots\ldots Z_n$, the maximum log likelihood estimate of Hurst h is given by:

$$\log l_{max}(h) = -\frac{1}{2}\log|C_n(h)| - \frac{n}{2}\log\frac{S(\bar{z},h)}{n} \qquad \ldots(73)$$

where $\bar{z}$ = mean of the series

$|C_n(h)|$ = Determinant of the correlation matrix given by $\left[\rho_{|i-j|}\right]$ as derived from equation (69).

$$S(\bar{z}, h) = (\underline{Z} - \bar{Z})^T \, |C_n(h)|^{-1} \, |(\underline{Z} - \bar{Z})|$$

$\underline{Z}^T = (Z_1, Z_2 \ldots\ldots Z_n)$ a 1x n vector.

The maximum likelihood estimate are calculated and the value of h which maximise the equation (73) is estimated by inverse quadratic interpolation technique (Kowalik and Osborne, 1968), Kumar (1980) has analysed the annual flows of the river Krishna and the Godavari river data to compare the performance of various estimates of h. It was observed that the MLE is the most consistent estimator of h. It was also observed, that both k and h* overestimate h if it is less than 0.7 and under estimate when h> 0.7. It is therefore recommended that Hurst h be calculated from the historical flows using the maximum likelihood approach.

7.6 Broken Line Models

Mejia et al (1972) presented an alternative long memory model, the broken line process. A simple broken line process is derived from the linear interpolation between the uniformly spaced independent variates with zero means. When two or more such processes each with a pre-assigned variance and starting points are added, a broken line process is obtained.

A simple broken line process is given by

$$\beta_i(t) = \{ \eta_m + \frac{\eta_{m+1} - \eta_m}{a} (t-ma) I \} \qquad \ldots(74)$$

where,

η_m = independently indentically distributed random number of zero mean and variance σ^2.

a = time distance among η_m

I = 1) where $ma < t < (m+1) a$
)
 0) otherwise

The variance of the fractions of the process is $(2/3\ \sigma^2)$ and the autocorrelation functions as

$$\rho_k = 1 - \frac{3}{4} (k/a)^2 (2 - \frac{k}{a}) \qquad 0 < k < a \qquad \ldots(75)$$

$$\rho_k = \frac{1}{4} (2 - \frac{k}{a})^3 \qquad k > a;\ \rho_k = 0 \quad a < k/2$$

For modelling, a BL process is formed by adding a finite number of broken line $B_i(t)$

$$a_t = \sum_{i=1}^{n} B_i(t) \qquad \ldots(76)$$

Mejia (1972) have related the parameters n, a_i, σ_i , i=1,2.....n with the mean, standard deviation and the correlogram structure of the process. The biggest disadvantage of the use of broken line model is very large number of parameters needed to model it. FGN models have therefore been preferred compared to BL models for long memory modelling.

7.7 Other Models

The classical ARMA models are based on the correlogram structure observed in the historical data, it does not account for Hurst h. Conversely, long memory models are based only on the Hurst h and not on the observed correlogram. Many researchers have therefore tried to formulate mixed models which account for both the parameters. Some important models in this class are (i) ARMA models with preassigned parameter values (O'Connell , 1971, 74) (ii) ARMA - Markov model (Lattenmaier and Burgess, 1977) (iii) White-Markov model Sen,1977).

7.8 ARMA Model (O'Connell, 1974)

O'Connell (1971, 74) showed that the process generated by an ARMA (1,1) model where ϕ is nearly one is similar to those by the FGN process.

$$\frac{Z_t - \bar{Z}}{\sigma_Z} = \phi_1 \frac{Z_{t-1} - \bar{Z}}{\sigma_Z} + a_t - \theta_1 a_{t-1} \qquad \ldots(77)$$

a_t is a normal independent process having a mean equal to 0 and variance $(1 - \phi_1^2) / (1 + \theta_1^2 - 2\phi_1\theta_1)$. It can be noticed from equation (77) that if ϕ_1, approaches unity, then the decay of the correlogram is very gradual, thereby increasing the memory of the process.

$$\rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 \quad \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \qquad \ldots(78)$$

$$\rho_2 = \phi_1 \rho_1$$

The estimation of parameters of ARMA (1,1) model which satisfies the Hurst h as well the ρ_1 can be easily estimated as follows:

1. Estimate $\bar{Z}$, σ_Z ,ρ_1 and Hurst h from a sequence of n observations
2. From the tables prepared by O'Connell (1974) identify the values of ϕ and θ , such that $E(h)_n = h$, and $E(\rho_1)_n = \rho_1$.
3. Knowing ϕ and θ , use equation (77) to model the flows.

The advantage of such a model is that they are very simple to use and the parameters ϕ and θ , can be estimated such that both Hurst h as well as the ρ_1 are satisfied. The disadvantages are that the model still follows $n^{0.5}$ law at large n and therefore does not model the Hurst phenomenon. Rodriguez-Iturbe (1974) showed this approximation may distort high frequencies properties meaning thereby that this model may not be able to produce realistic crossing properties.

7.9 ARMA Markov Models

Lattenmaier and Burges (1977) proposed a mixed model called ARMA-Markov model which uses the Hurst coefficient h as an explicit parameter. The model consists of five parameters (i) The Markov and ARMA variance fractions, C_1 and C_2 respectively (ii) The Markov and ARMA lag one autocorrelation coefficients ρ_m and ρ_{am} respectively, and (iii) the autoregressive parameter ϕ of an ARMA process. The moving average parameter θ of the ARMA process is uniquely defined by ϕ and ρ_{am}.

The modelling equation of zero mean and unit variance ARMA - Markov process is

$$Z_t = \rho_m Z_{t-1}^{(m)} + a_t^{(m)} + \phi_1 Z_{t-1}^{(am)} - \theta_1 a_{t-1}^{(am)} + a_t^{(am)} \qquad \ldots(79)$$

in which $a^{(m)}$ and $a^{(am)}$ are the independent process having zero mean and variance $C_1(1 - \rho_m^2)$ and $C_2\{(1-\phi_1^2) / (1+\theta_1^2 - 2\phi_1\theta_1)\}$respectively.

The autocorrelation function of the process is fitted to the theoretical autocorrelation function of the FGN process at three specified lags k_1, k_2 and k_3. The lag one autocorrelation coefficient may be arbitrarily specified. The parameters of the models are obtained by solving the following system of equations:

$$C_1 + C_2 = 1.0$$

$$C_1\rho_m + C_2\rho_{am} = \rho_1$$

$$\begin{aligned} C_1\rho_{m,k_1} + C_2\rho_{am,k_1} &= \rho_{k_1} \\ C_1\rho_{m,k_2} + C_2\rho_{am,k_2} &= \rho_{k_2} \\ C_1\rho_{m,k_3} + C_2\rho_{am,k_3} &= \rho_{k_3} \end{aligned} \qquad \ldots(80)$$

in which C_1, C_2, ρ_m, ρ_{am}, ϕ_1, are constrained to be between 0 and 1, ρ_1 = the desired lag one (acf) and ρ_{am} is the autocorrelation of the FGN process

$$\rho_k = \frac{1}{2}\{(k+1)^{2h} - 2k^{2h} + (k-1)^{2h}\} \qquad \ldots(81)$$

Lattenmaier and Burgess (1977) suggest to take the values k_1, k_2 and k_3 to be approximately n/8, n/2 and n in which n is the length of the time series to be modelled.

Equation (80) may be solved by using Newton's method to give model parameters for given ρ_1 and h. The second parameter θ of the ARMA process is obtained from

$$\rho_{am} = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2 - 2\phi_1\theta_1} \qquad \ldots(82)$$

If $a^{(m)}$ and $a^{(am)}$ are from a normal process then the Z_t series will be normal. To model the necessary skewness C_{sz}, $a^{(m)}$ and $a^{(am)}$ should have the following skewness C_{sa} in the Wilson Hilferty transformation

$$C_{sa} = \frac{C_{sz}}{\left[\frac{C_1(1-\rho_m^2)^{3/2}}{1-\rho_m^3} + \left[\frac{1-\phi_1^3+3\phi_1\theta_1^2-2\phi_1^2\theta_1}{1-\phi_1^3}\right]\left[\frac{C_2(1-\phi_1^2)}{1+\theta_1^2-2\phi_1\theta_1}\right]\right]^{3/2}}$$

$\ldots(83)$

7.10 White Markov Process

Sen (1977) proposed a linear stationary process called white-Markov process (wM_p) to model the Hurst phenomenon. The model is made up of a first order Markov process and a white noise process which is independent of the Markov process. The standard (wM_p) having zero mean and a unit variance can be written as

$$Z_t = \alpha^{\frac{1}{2}} Z_t^{(m)} + (1-\alpha)^{\frac{1}{2}} a_t \qquad ..84)$$

in which $Z_t^{(m)}$ is a Markov process with zero mean and unit variance, a_t is a white noise process with zero mean and unit variance and α is a constant satisfying $0< \alpha < 1.0$.

The lag k autocorrelation of the process is given as

$$\rho_k = \alpha\rho_m^{|k|} \qquad k > 1 \qquad ...(85)$$

in which ρ_m is the autocorrelation of the Markov process.

Sen (1977) carried out a series of Monte Carlo experiments and presented a set of tables (similar to those of O'Connell, 1974) to obtain α and ρ_m for a given pair of h and ρ_1. A major drawback in this procedure is its inability to preserve the population h as an explicit model parameter.

To model skewed flows, the random number a_t should have the following skewness:

$$C_{sa} = \frac{C_{sz}}{\left\{\alpha \frac{(1-\rho_m^2)^{3/2}}{1-\rho_m^3} + (1-\alpha)^{3/2}\right.} \qquad ...(86)$$

8.0 MULTISITE LONG MEMORY MODELS

The development of long memory models in the multisite domain is at an early stage, the only known work is that of Matalas and Wallis (1971 a) for the fractional noise process, Mejia et al (1974) for the broken line process, O'Connell (1974) for the ARMA models and Weiss (1977) for the shot noise models. However, all these models have simply been made operational by the respective researchers for their preferred class of models. Not much work has been done on the comparative use of Brownian and fractional Brownian motion model in multisite modelling. Similarly, the properties and behaviour of different models in maintaining long run serial and cross correlation matrices using multisite long memory models are yet to be examined. In fact, the whole area of multisite long memory modelling remain virtually unexplored.

9.0 FINAL COMMENTS

A short review on hydrologic time series with particular emphasis on river flow time series modelling has been presented. Time series modelling for real time forecasting and the conjoint rainfall runoff time series modelling has not been attempted. A separate review paper may be required to cover these aspects.

Areas on which further work is required is briefly detailed as follows:

i) With respect to short memory models, further work is required to estimate the effect of skewness present in the river flow series on its different properties. It is well known that skewness has a very significant effect on the crossing properties like run and range analysis. For example, Kumar (1980) showed that positive skewness increases (decreases) the expected surplus (deficit)run sum whereas, it has negligible effect on the estimation of run length etc. Further work is required to mathematically formulate the effect of skewness on crossing properties.

ii) Not much effort has been put by hydrologists in identifying a most suitable model. Probably, more attention is required. In this connection, it may be mentioned that extensive work has been done by the Control Engineers. The Control Engineering Journals have numerous papers in the selection of parsimonious model. The Aikike information criterion and the posterior probability criterion suggested by them looks quite promising. It is suggested that more of such criterions be used in hydrologic time series analysis.

iii) Daily flow modelling as suggested by Weiss (1977) opens an interesting vista for realistic simulation of ascension- recession behaviour of daily hydrographs. This method needs further development especially in reducing the number of parameters to be calculated from the historical records. To the author's knowledge, no particular study using shot noise model has been made for Indian catchment. The experience of using such a model for Indian rivers would be highly informative.

iv) Though extensive work has been done on monthly flow modelling, the use of such models for modelling monthly/pentad flows for Indian rivers may require some modification. This is because many Indian rivers have nearly zero flows during certain parts of the year. This will require incorporating definite probability for zero flows in the model. In this connection, further works needs to be done on the lines of models developed for ephemeral streams by Yakowitz (1973), Srikanthan Mcmahon (1980).

v) Differential persistence is another area where not much work has been reported. Pattern recognition technique (Pannu and Unny, 1980) may lead to better insight on the run of high and low flows. Further work on the modelling of differential persistence will be highly welcome.

vi a) Multisite modelling tend to be tedious as the number of sites increases and the parameter estimation problem increases when higher lag serial and corss-correlation matrices are to be maintained. Decoupled multisite models as suggested by Ramaseshan (1975) make multisite modelling less tedious. However, further work is needed on the sampling properties of the correlational matrices whose historical preservation is required. It is also to be seen how the decoupled multisite model reproduces in each series low and high frequency values which are in correct chronological relation.

Vi b) Recently, Johnson and Bras (1980) have developed a stochastic model for short term (of the order of one hour or less) rainfall prediction. The model includes velocity and direction of storm movement as explicit parameters. The use of this methodology on Indian catchments will be highly informative.

vii) Long memory models in hydrology are controversial. The major argument against the short memory models is that it does not model the Hurst h with the result it can not model the distributions of deficit and duration of extreme high and low flows observed in long historical record. On the other hand, the argument against long memory model is that it is only a mathematical exercise without any physical meaning. It is generally felt that the use of short memory models leads to severe under estimation of storage capacities. However, Klemes (1981) has shown that the differences in reservoir performance reliability obtained on the basis of long and short memory models is small compared to (i) the accuracy of measurement of the socio-economic impact of reliability changes and(ii) the accuracy of estimating the reliability itself on the basis of available stream flow records. Similarly, it is felt that long memory models are the only known models that can retain the rescaled adjusted range (r.a.r.) Again a study by Hipel (1977) demonstrated from the analysis of 23 time series that a properly validated ARMA model do preserve the r.a.r. These studies pose a question - are long memory models superfluous? This question can be answered only if an in depth study of long and short memory models is done and the various crossing properties derived by these models compared. This will be a significant contribution to hydrologic literature.

--- . ---

REFERENCES

1. Adamowksi, K. (1971), 'Spectral density of a river time flow series', J.Hydrol., Vol.14, pp 43-52.

2. Aikeke, H. (1974), 'A new look at the statistical model identification IEEE Trans.' Automatic Control, Vol.19, pp 716-723.

3. Andel, J., and Balek, J.,(1971), 'Analysis of periodicity in hydrological sequence', J. Hydrol, Vol 14, pp 66-82.

4. Anderson, R.L., (1942), 'Distribution of serial correlation coefficient', Annals of Mathematical Statistics, Vol 13, pp.1-13.

5. Anis, A.A. and Lloyd, E.H. (1953), 'On the range of partial sums of a finite number of independent normal variates', Biometrika, Vol.40, pp 35-42.

6. Anis, A.A. and Lloyd, E.H. (1975), 'Skewed input and the Hurst effect', Jnl. of Hydrol., Vol 26, pp 39-53.

7. Anis, A.A. and Lloyd, E (1976), 'The expected value of the adjusted rescaled Hurst range of independent normal deviates',Biometrika, Vol 63, pp 111-116.

8. Bayazit, M.(1982),'Three state Markov model for differential persistence', Jnl. of Hydrol.,Vol 55, pp 339-346.

9. Beard, L.R. (1965),'Use of interrelated records to simulate streamflows', Jnl.Hyd.Division, ASCE, Vol 91, pp.13-22.

10. Beard, L.R. (1965 b),'Simulation of daily stream flows', Intl. Symp. in Hydrology, Fort Collins, Colorado.

11. Bendat, J.S. and Piersol, A.G. (1971), 'Random data analysis and measurement procedures', Wiley, New York.

12. Bernier, J.(1970), 'Probabilistic models for hydrological synthesis', Symp.Math Models in Hydrology, Warsaw, Vol 1, pp.333-342.

13. Box, G.E.P. and Jenkins, G.M.(1970),'Time series analysis, forecasting and control', Hoden Dey Publishing Co., California.

14. Box, G.E.P. and Cox, D.R. (1964),'Analysis of transformation' Jnl.Royal Statistical Soci. Vol.28, pp.211-252.

15. Brier, G.W. (1981), 'Some statistical aspects of long term fluctuation in solar and atmospheric phenomenon', Annals New York Academic Sciences, Vol 95, pp173-187.

16. Brown, R.G. (1963), 'Smoothing, forecasting and prediction of discrete series', Prentice Hall, Englewood, Cliffs,New Jersey.

17. Buish and T.E. (1977),'Stochastic modelling of daily rainfall sequences', Meddelingen Landbouewhogeschcool Wageningen,Netherlands, Vol.77.

18. Burges, S.J. (1972),'Some problems with lognormal Markov runoff models', J.Hyd.Division, ASCE, Vol.98, pp 1487-1495.

19. Carlson, R.E., Mac Cormick, A.J.A. and Watts, D.G.(1970),'Application of Linear random models to four annual stream flow series', Water Resources Research, Vol 6, pp 1070-78.

20. Chander, S., Spolia, S.K. and Kumar, A.(1979),'Flood stage forecasting in rivers using stochastic modelling', IRIA, France,pp 5.15.

21. Chander, S., Kumar, A. and Goyal, S.K.(1979),'Stochastic models for monthly rainfall forecasting and synthetic generation-comments', Jnl. of App. Meteorology, American Meteorological Society, Vol 18, pp 1381.

22. Chander, S., Spolia, S.K., and Kumar, A. (1980),'Autorun Analysis of hydrologic time series - A comment', Jnl. of Hydrology, Vol. 45, pp.33-339.

23. Chander,S., Kumar, A., Goyal, S.K., Spolia, S.K. and Kapoor, P.N. (1980),'Choice of transformation in seasonal Box-Jenkins modelling for hydrological forecasting', Intl.Symp. on Hydrological forecasting, Oxford, IAHS, No.129.

24. Chatfield, C., and Prothero, D.L.(1973),'Box Jenkins seasonal forecasting problems in a case study', Jnl.Royal Statistical Society, Ser A, pp 295-35.

25 Chi, M., Keal, E and Young, G.K. (1977),'Practical application of fractional Brownian motion and noise to synthetic hydrology', Water Resources Research, Vol.9, pp.1569-1582.

26. Chow, V.T. and Ramaseshan, S. (1965),'Sequential generation of rainfall runoff data', Jnl. of Hyd. Division, ASCE, Vol 91,pp 205-220.

27. Chow, V.T. and Kareliotis, S.G. (1970),'Analysis of stochastic hydrologic systems', Water Resources Research, Vol.6, pp.1569-1582.

28. Clarke, R.T. (1973),'Mathematical models in Hydrology', Irrigation and Drainage paper, No.19,FAO.

29. Claveland, W.S.,(1972), 'The inverse autocorrelation of a time series and their application', Technometrics, Vol.14,pp.277-288.

30. Cline,T.B. (1979), 'Selecting seasonal streamflow models', Water Resources Research, pp.975-984.

31. Codner, G.P. and McMahon, T.A.(1973),'Log normal streamflows generation model re-examined', Jnl.Hyd.Division, ASCE,Vol 99, pp.1421-1431.

32. Crosby, D.S. and Maddock, T, III (1970),'Estimating coefficient of a flow generator from monotype sample of data', Water Resources Research, Vol 6, pp.1079-1086.

33. Davies, R.K.(1968),'Range of choice in water management', Baltimore, John Hopkins Press.

34. Delleur, J.W.,Tao, P.C. and Kanvaas, M.L.(1976),'An evaluation of practicability and complexity of some rainfall runoff time series models', Water Resources Research, Vol 12, pp.953-970.

35. Delleur, J.W. and Kanvaas, M.L.(1978),'Stochastic models for monthly rainfall forecasting and synthetic generation', Jnl.App. Meteorology,' Vol 17, pp.1528-1536.

36. Dixey.

37. Domokos, M.(1970),'Estimation of distribution and density function and their application in the management of water resources' Acta Technica Acad.Services Hundary, Vol 69, pp.381-413.

38. Dooge, J.C.I. (1972),'Mathematical models of hydrological systems' Modelling of Water Resources Systems (Ed.A.K. Biswas),Vol 7, pp. 170-188.

39. Durbin, J.(1962),'Trend elimination by moving arrange and variance difference filters', Bull. Intl.Stat.Inst.,Vol 39, pp. 131-141.

40. Feller, W. (1951),'The asymptotic distribution of range of sum of independent random variable', Annals of Mathematical Statistics Vol.22, pp.427-432.

41. Fiering, M.B. (1964),'Multivariate technique for synthetic Hydrology', Jnl.Hyd.Div.,ASCE,Vol 90,pp.43-60.

42. Fiering, M.B. (1967),'Streamflow synthesis', Cambridge,Harvard Univ.Press.

43. Fiering, M.B.(1968),'Schemes of handling inconsistent matrices', Water Resources Research, Vol.4,pp.291-297.

44. Fiering, M.B. and Jackson, B.B.(1971),'Synthetic streamflows', American Geophysical Union, Water Resources Bulletin, Monograph 1.

45. Frost, J. and Clarke, R.T. (1973),'Use of cross-correlation between hydrological time series to improve estimates of lag one auto-correlation parameters', Water Resources Research Vol 9,pp.906-917.

46. Gupta, V.K. and Fordham, J.W. (1972),'Streamflow synthesis - A case study', Jnl.Hyd.Division, ASCE, Vol 98,pp.1049-1055.

47. Hall, W.A., Askew, J.J. and Yeh, W.W.G.(1969),'Use of critical periods in reservoir analysis'Water Resources Research, Vol.5, pp. 1205-1215.

48. Hamlin, M.J. and Kottegoda, N.T. (1971),'Extending the record of the Teme'; Jnl.Hydrol. Vol.12, pp.110-116.

49. Harrison, P.J. and Stevens, C.F. (1976), 'Bayesian Forecasting', Jnl. Royal Statistical Society, Ser. B.,Vol.38, pp.205-247.

50. Healy, F (1969), 'An AS-6 Algorithm', Applied Statistics.

51. Hipel, K.W., Mcleod, A.I. and Lennox, W.C. (1977),'Advances in Box Jenkins modelling', Water Resources Research, Vol.13, pp.567-575.

52. Holloway, J.L. (1958), 'Smoothing and filtering of time series and space fields', Advances in Geophysics, Vol.4, pp.351-389.

53. Hunfschmidt, M.M. and Fiering, M.B. (1966),'Simulation technique for design of water resources systems', Harvard University Press.

54. Hurst, H.E. (1951), 'Long term storage capacities of reservoirs Transanctions', ASCE, Vol.116, pp.770-808 .

55. Hurst, H.E. (1957),'A suggested statistical model of some time series that occur in nature', Nature, Vol.180, pp.494.

56. Jackson, B.B. (1975),'The use of streamflow models in planning', Water Resources Research, Vol.11, pp.54-63.

57. Jackson, B.B. (1975 b), 'Use of Birth-death models for streamflow modelling', Water Resources Research, Vol.11, pp.64.

58. Jenkins, R.H. and Watts, D.G. (1968), 'Spectral analysis and its applications', Holden Dey, San Francisco, USA.

59. Johnson, R.E. and Bras, R.L. (1980), 'Multivariate short term rainfall prediction', Water Resources Research, Vol.16, pp.173-185.

60. Jones, R.H. and Breslford, W.M. (1967),'Time series with periodic component', Biometrika, Vol.54, pp.403-408.

61. Julian, P.R. (1967),'Variance spectrum analysis', Water Resources Research, Vol.3, pp.831-845.

62. Kashyap, R.L. and Rao, A.R. (1976),'Dynamic stochastic model from empirical data', Academic Press, New York.

63. Kashyap, R.L. (1980),'Inconsistency of the AIC rule for estimating the order of the autoregressive model', Tech. Report Deptt. of Elect. Engg., Purdue University.

64 Kendall, M.G.(1964),'Time series analysis', Charles Griffin and Co. London.

65. Kendal, M.G. and Stuart, A (1973),'The Advanced Theory of Statistics', Vol.2., Griffin and Co, London.

66. Kendall, M.G. and Stuart, A (1976),'Advanced theory of statistics, Vol.III,Griffin and Co, London.

67. Kindler, J and Zuberek, W (1976),'On some multisite multiseason streamflow generation model' Intl. Inst. App. System Analysis, Luxemberg, Research Memo. No.76.

68. Kirby, W (1972), 'Computer oriented Wilson-Hilferty transformation that preserves the first three moments and the lower bound of the Pearson type III distribution', Water Resources Research, Vol.8, pp.1251-1254.

69 Kirby, W (1974),'Algebraic boundedness of sample statistics', Water Resources Research, Vol.10,pp.220-222.

70. Kisiel, C.C.(1969),'Time series analysis of hydrologic data', Advances in Hydrosciences (Ed.Chow), Chap.5, pp.1-119.

71. Klemes, V (1974),'The Hurst Phenomenon - A puzzle', Water Resources Research, Vol.10,pp.675-688.

72. Klemes, V(1979),'Unreliability of reliability estimates of storage reservoirs performance based on short streamflow records', Reliability in Water Resources, WRP, Fort Collins, pp.193-205.

73. Klemes, V.(1981),'Long memory flow models in reservoir analysis. What is their practical value?', Water Resources Research, Vol.17,pp.737-751.

74. Kottegoda, N.T. (1970),'Statistical methods of river flow synthesis for water resources assessment', Proc. Inst. of Civil Engrs. London.

75. Kottegoda, N.T. (1972),'Flood evaluation- can stochastic models provide an answer?', Proc. Intl.Symp. on Uncertainties in Hydrology and Water Resources, Univ. of Arizona, pp.105-114.

76. Kottegoda, N.T.(1973),'Flood estimation by some data generation technique', Floods and Drought Proceedings, WRP, Fort Collins, pp. 189-199.

77. Kottegoda, N.T. and Elgy, J. (1977),'Infilling missing data', Proc. Intl. Symp. on Hydrology, Fort Collins, Colorado.

78. Kottegoda, N.T.(1979),'Stochastic Water Resources Technology', John Wiley and Sons, New York.

79. Kowalik, J. and Osborne, M.R. (1968),'Methods for unconstrained optimisation problem', American Elsevier, New York.

80. Kumar, A.(1980), 'Prediction and Real time hydrological forecasting', Ph.D. thesis, IIT Delhi.

81. Kumar, A. and Devi, R.(1982),'Stochastic generation of annual streamflows - Discussion', Jnl.of Hyd.Division, ASCE,HYI,pp. 255-258.

82. Lawrence, A.J. (1976),'A reconsideration of the Fiering two station model', Jnl. of Hydrology, Vol.29,pp.77-85.

83. Lawrence, A.J. and Kottegoda, N.T.(1977),'Stochastic modelling of river flow time series', Jnl.Royal Statistical Society, Ser.A., Vol.140,pp.1-47.

84. Ledolter, J. (1978a),'A general class of stochastic models for hydrological sequences', Jnl. of Hydrology, Vol.36,pp.309-325.

85. Ledolter, J.(1978b),'The analysis of multivariate time series applied to problems in hydrology', Jnl of Hydrology, Vol.36,pp. 327-352.

86. Lin, Y.W.(1960),'Statistical Theory of communication', Wiley, New York.

87. Lenton, L.L.,Rodriguez-Iturbe, I. and Schaake, Jr. J.C. (1974), 'The estimation of in the first order auto-regressive model - Bayesian approach', Water Resources Research, Vol.10, pp.227-241.

88. Lettenmaier, D.P. and Burges, S.J.(1977),'Operational assessment of hydrologic models for long term persistence', Water Resources Research, Vol.13, pp.113-124.

89. Macmicheal, F.C. and Hunter, J.S.(1972),'Stochastic modelling of temperature and flows in rivers', Water Resources Research, Vol.8, pp.87-88.

90. Maissis, A.H.(1977),'Optimal filtering technique for hydrological forecasting', Water Resources Research, Vol. pp.87-98.

91. Mandelbrot, B.B. and Wallis, J.R.(1968),'Noah, Joseph and operational hydrology', Water Resources Research, Vol.4,pp.909-918.

92. Mandelbrot,B.B. and Wallis, J.R.(1968a),'Computer experiments with fractional Gaussian noise', Water Resources Research, Vol. 5, pp.228-267.

93. Mandelbrot, B.B. and Wallis, J.R.(1969b) , 'Robustness of rescaled range R/s in the measurement of noncyclic long run statistical dependence', Water Resources Research, Vol.5, pp.967-968.

94. M. ndelbrot, B.B. and Wallis, J.R.(1969c),'Some long run properties of geophysical records', Water Resources Research, Vol.5,pp.321-340.

95. Mandelbrot, B.B. (1976),'Note on the definition and stationarity of fractional Gaussian noise', Jnl. of Hydrology, Vol.30,pp.407-409.

96. Markovic, R.D.(1965),'Probability function of best fit distribution of annual precipitation and runoff', Hydrology paper no.8,Colorado State Univ. Colorado.

97. Matalas, N.C.(1967),'Mathematical assessment of synthetic hydrology' Water Resources Research, Vol.3,pp.937-945.

98. Matalas, N.C. and Huzzen, C.S.(1967),'A property on the range of partial sums', Intl. Hydrology.Symp. Colorado State Univ. Colorado.

99. Matalas, N.C. and Wallis, J.R.(1971),'Statistical properties of multivariate fractional noise process', Water Resources Research Vol.9,pp.1271-1285.

100. Matalas, N.C. and Wallis, J.R.(1971b),'Correlation constraints for generating processes', Intl. Symp. on Mathematical modelling Warsaw (IAHS publ.100), pp.372-376.

101. Matalas, N.C. and Wallis, J.R.(1973),'Eureka; it fits a Pearson type III distribution', Water Resources Research, Vol.9, pp.281-289.

102. Mcleod, A.F. and Hipel, K.W. (1978),'Perservation of rescaled adjusted range- A reassessment of Hurst Phenomenon', Water Resources Research, Vol.19,pp.491-508.

103. Mcleod, A.I.Hipel, K.W. and Lennox, W.C.(1977),'Advances in Box-Jenkins modelling, Part 2 Applications', Water Resources Research, Vol.17,pp.577-586.

104. Mcleod, A.I. and Hipel, K.W.(1978),'Simulation procedure for B.J. model',Water Resources Research, Vol.19.pp 969-975.

105. Mcmahon, T.A. and Miller,A.J. (1971),'Application of the Thomas Fiering model to skewed hydrologic data', Water Resources Research, Vol.7,pp.1338-1339.

106. Megerian, E and Pentland, R.L. (1968),'Simulation of Great Lake Basin Water Supplies',Water Resources Research, Vol.4,pp.11-17.

107. Mejia, J.M. and Rodriguez-Iturbe, I.R.(1974),'Correlation links between normal and log-normal process', Water Resources Research, Vol.10,pp.689-690.

108. Mejia, J.M. and Rouselle, J (1976),'Disaggregation models in hydrology revisited', Water Resources Research, Vol;12,pp.185-186.

109. Merkrecher, A.I. and Delleur, J.W.(1974),'Application of seasonal parametric linear stochastic models to monthly flow data', Water Resources Research, Vol.10,pp.246-255.

110. Moran, P.A.P.(1957),'The statistical treatment of flood flows', Trans. American Geophys. Union, Vol.38,pp.519-523.

111. Moran, P.A.P. (1970),'Simulation and evaluation of complex water system operation', Water Resources Research, Vol.6, pp.1737-1742.

112. Moss, M.E.(1972)',Reduction of uncertainities in autocorrelation by the use of physical models', Proc. Intl. Symp. on Uncertainities in Hydrologic and Water Resources System, Univ. of Arizona, pp. 203-209.

113. O'Connell, P.E.(1971),'A simple stochastic modelling of Hurst Law', Math. Models in Hydrology Symp.,Warsaw, Vol.1,pp.169-187.

114. O'Connell, P.E.(1977),'Discussion on stochastic modelling of river flow time series', Jnl.Royal Stat. Society, Vol.40,pp.32.

115. Ozaki, T.(1980),'On a model of non-linear feed back system for river flow prediction', Water Resources Research, Vol.16,pp.225-231.

116. Pannu, U.S. and Unny, T.E.(1980),'Stochastic synthesis of hydrologic data based on the concept of pattern recognition', Jnl. of Hydrology, Vol.46,pp.5-34.

117. Pentland, R.L. and Cuthbert, D.R.(1973),'Multisite data flow generator', Water Resources Research, Vol.9, pp.470-473.

118. Phien, H (1979),'Streamflow simulation - A model based on Canonical expansion- comments', Jnl. of Hydrology, pp.383-384.

119. Quimpo, R.G.(1967),'Stochastic model for daily riverflow', Sequences Hydrology paper 18, Colorado State University.

120. Quimpo, R.G.(1968),'Stochastic analysis of daily flow', Jnl. of Hyd. Division, ASCE, Vol.94,pp.43-57.

121. Ramaseshan, S. and Krishnaswami, M.(1975),'Decoupled stochastic models for multisite stream flows', Prof. Second World Congress, IWRA, New Delhi,pp.45-60.

122. Ramaseshan, S. (1976),'Time series analysis of hydrologic data. Lecture notes on advanced course on system analysis of hydrological problems', IIT,Kanpur.

123. Rodriguez-Iturbe, I., Mejia, J.M. and Dawdy, D.R.(1972),'Streamflow simulation. A new look at Markovian models, fractional Gaussian noise and crossing theory', Water Resources Research,Vol.8,pp.921-930.

124. Roesner, L.A. and Yevjevich, V.(1966),'Mathematical models for time series of monthly precipitation and monthly runoff', Hydrology paper no.15, Colorado State Univ. Fort Collins.

125. Schaake, Jr., J.C. and Fiering, M.B.(1967),'Simulation of a national flood insurance plan', Water Resources Research, Vol.3, pp.913-930.

126. Sen, Z.(1977),'The White Markov process for modelling Hurst Law', Proc. Third Intl.Symp. in Hydrology, Fort Collins.

127. Sen,Z.(1978),'Autorun analysis of hydrologic time series', Jnl. of Hydrology, Vl.36, pp.75-85.

128. Singh, K.P. and Lonnquist, C.G. (1974),'Two distribution method of modelling and sequential generation of monthly streamflows', Water Resources Research, Vol.10,pp.763-773.

129. Slack, J.R. (1973)', Self denial by conditional model', Water Resources Research, Vol.19, pp.247-249.

130. Slutzky, E.(1937),'The simulation of random causes as the source of cyclic processes', Econometrika, Vol.5, pp.105-146.

131. Smirnov, N.P.(1969),'Causes of long period streamflow fluctuations', Amer.Geophys.Union, Soviet Hydrology, Selected paper 3, pp.308-314.

132. Solari, M.E. and Anis, A.A.(1957),'The mean and the variance of the maximum of the adjusted partial sums of a finite number of independent random deviates', Annals of Mathematical Statistics, Vol.28, pp.706-716.

133. Spolia, S.K. and Chander, S.(1974),'Modelling of surface runoff system by an ARMA model', Jnl. of Hydrol, Vol.22,pp.317-372.

134. Spolia, S.K. and Chander, S.(1977),'Streamflow simulation - A model based on cannonical expansion', Jnl. of Hydrol. Vol.35, pp.279-288.

135. Srikanthan, R. and Mcmahon, T.A.(1979),'Generation of annual streamflows using the fractional Gaussian noise model', Jnl. of Hydrol, Vol.37, pp.13-21.

136. Srikanthan, R. and Mcmahon, T.A.(1980),'Stochastic generation of monthly flows of ephemeral streams',Jnl.of Hydrol.,Vol.47,pp.19-40.

137. Srikanthan, R. and Mcmahon, T.A.(1980),'Stochastic generation of annual streamflows', Jnl. of Hyd.Division, ASCE,pp.2010-2027.

138. Tao, P.C. and Delleur, J.W.(1976),'Seasonal and non-seasonal ARMA models in hydrology', Jnl. Hyd. Division, ASCE,pp.1541-1559.

139. Thomas, H.A. and Fiering, M.B. (1962),'Mathematical analysis of streamflow sequences for the analysis of river basin by simulation', Design of Water Resources System,(Ed.Mass et al) Cambridge University Press, Mass,pp.459-493.

140. Valencia, R.D. and Schaake, J.C.(1973),'Disaggregation process in stochastic hydrology',Water Resources Research, Vol.9,pp.580-585

141. Wallis, J.R. and Matalas, N.C.(1970),'Small sample properties of H and K estimators of Hurst coefficient h', Water Resources Research,Vol.6,pp.1583-1594.

142. Wallis, J.R. (1972),'Small sample estimates of correlation coefficient',Water Resources Bulletin.

143. Weiss, G.(1977),'Shot noise modells for synthetic generation of multisite daily streamflow data', Water Resources Research, Vol.13,pp.101-108.

144. Weiss, G.(1977),'Stochastic modelling of riverflow time series comments', Jour. Royal Statist.Society, Ser A, Vol.140, pp.41.

145. Whitakar, J.(1973),'A note on the generation of gamma random variables with non-integer shape parameters ', Proc.2nd Intl. Symp.Hydrology, Fort Collins, Colorado.

146. Wilson, E.B. and Hilferty, M.M.(1931).'Distribution of Chi squares',Proc. Nat. Acad. Sciences, USA, pp.684-688.

147. Yakowitz, S.J.(1973),'A stochastic model for daily streamflow in an arid region', Water Resources Research, Vol.9,pp.1271-1285.

148. Yevjevich, V.(1972),'Stochastic processes in hydrology', Water Resources Publication, Fort Collins, Colorado.

149. Yevjevich, V.(1972),'Structural analysis of hydrologic time series', Hydrology paper 56, Colorado State University, Fort Collins.

150. Yevjevich, V.(1973),'Determinism and stochasticity in hydrology', Jnl. of Hydrology, Vol.22,pp.225-238.

151. Young, G.K. and Pisano, W.C.(1968),'Operational Hydrology using residuals',. Jnl. of Hyd. Dividion,ASCE, Vol.94,pp.909-923.

***