RN-18

TIME SERIES ANALYSIS MODELS

SATISH CHANDRA
DIRECTOR

STUDY GROUP

S M SETH
N K GOEL

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

1985-86

CONTENTS

LIST OF FIGURES

ABSTRACT

Time series analysis belongs to major statistical techniques used in the extraction of information on hydrologic and water resources random variables from observed data. This report gives a brief review on time series models and steps used for time series modelling. Various criteria for the classification of time series models are presented and described. Available time series models are explained in the light of short memory models and long memory models. Short memory models include autoregressive (AR), moving average (MA), autoregressive moving average(ARMA), and autoregressive integrated moving average (ARIMA) models. Long memory models such as fast fractional Gaussian noise, filtered fractional Gaussian noise and broken line models are then described. Generation of daily data by shot noise model has been given. In the end disaggregation model and multisite models have been explained.

Some of the areas in which further study and research are needed have been identified by the review of literature. These include (i) time series analysis of water quality and quantity to meet the solution of complex environmental problems, (ii) development of more comprehensive families of time series models, (iii) physically based time series models, (iv) development of daily flow generating models

with lesser parameters, (v) differential persistence and (vi), application of time series models (after modification) to Indian rivers as many of them have nearly zero flows during non-monsoon season (Nov.-May).

1.0 INTRODUCTION

A time series represents a set of observations that measure the variation in time of some dimension of a phenomenon such as precipitation, windspeed, river flow etc. The time series can be considered in continuous or discrete form. Most practical applications in hydrology consider the discrete form primarily as it is easier to handle the discrete time series on a digital computer.

Time series can be classified as stationary time series or nonstationary time series. If the expected value of statistical parameters does not change with time, the time series si said to be stationary, otherwise nonstationary.

Most of the statistical methods used in hydrologic studies assume that the observations are independently distributed in time. In other words, the occurrence of an event in assumed to be independent of all previous events. This assumption is not always valid for hydrologic time series e.g. observations of daily discharges do not change much from one day to the next and there is a tendency for the values to cluster as high values tend to follow high values and low values tend to follow low values. Thus, the daily discharges are not independently distributed in time. Though this is also true that this dependence between hydrologic observations decreases as time base increases.

1.1 Components of Hydrologic Time Series

Hydrologic time series is composed of two components i.e. (i) nonrandom element (deterministic component) and (ii) random element (stochastic component). A non-random element is said to exist when observations separated by K time units are dependent. If the values of x_i are linearly dependent upon the values of x_{i+k}, then the correlation between x_i and x_{i+k} may be taken as the measure of dependence. This correlation is referred to as the kth-order serial correlation, represented by γ_k.

If a time series is random, $\gamma_k = 0$ for all values of $k \geq 1$. However for a sample of finite size, computed values of γ_k may differ from zero because of sampling errors. Since N is small for most hydrological sequences the sampling errors are very large.

The non-random element (deterministic component) may be composed of both a trend (long term movement) and an oscillation about the trend (periodicity). Both of these need not be present in a particular time series. The first step in analysing a time series is to separate the non-random element from the random element.

The aim of time series analysis is to identify and separate deterministic component i.e. trend and periodicity from the original series. The resultant stochastic component is then modelled and the combined effect is determined by superimposing deterministic component on it. An excellent discussion on determinism and stochasticity in time series is given by Yevjevich (1973) and Kumar (1982). Analysis of deterministic and stochastic components is described in subsequent sections.

1.1.1 Deterministic Component Analysis

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 is generally 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 to the original series.

There are number of statistical tests to detect the presence of trend in a time series. Kendall's rank correlation test (Kendall and Stuart, 1973, Kottegoda,1979) and linear regression tests can be used to check whether the time series is trend free or not. The rank correlation test as well as the linear regression test are not valid for detecting the presence of non-homogeneties 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 .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 the 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 have been applied in meteorology (Holloway, 1958 and 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).

Periodic Component: There can be two types of periodic components i.e. (i) long range periodicity and (ii) short range periodicity, in a time series. These are being explained subsequently.

Long range periodicity:

A hydrologist is interested to know whether the series has a long range periodicity (more than a year). For example, it is commonly said that drought occurs once in five years. Many climatologists 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 sunspot number varies approximately in a long term periodic manner, the period ranging from 13 years to 8 years with a mean of 11.1 years. Similarly, Smirnov (1969) found significant correlation between sunspot and the mean flows in the Volga. Indian scientists at I.M.D., 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 USA and 16 series of annual runoff from all over the world with the sunspot number and they could not find any significant correlation. It can be assumed that the observed river flow series does not follow any long range periodic behaviour (Kumar, 1982).

Short range periodicity (seasonality): The river flow series may be obtained as an annual, monthly, ten daily, pentad (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. Seasonality is observed in pentad and daily data (Bernier, 1970) also.

Seasonality can be removed by 'prewhitening.' A common method of prewhitening is to standardize and remove periodicity. If $x_{t,\tau}$, $t = 1,2 \ldots$; and $\tau = 1,2,\ldots,12$ is a monthly series, then

$$z_{t,\tau} = \frac{x_{t,\tau} - \bar{x}_\tau}{\sigma_\tau}$$

is a standardized series. Here $\bar{x}_\tau$ and σ_τ are mean and standard deviation of the τth month. The drawback of standardization is that large number of parameters are required e.g. in case of daily data 365 values of mean and 365 values of standard deviation are required to standardize the series. In case of short samples, the estimate of such large values may lead to sampling errors. In such case, the values of $\bar{x}_\tau$ and σ_τ are smoothened by harmonic analysis. A good description of harmonic analyses is given by Roesner and Yevjevich (1966).

1.1.2 Stochastic component analysis

The remaining series i.e. series after removal of deterministic component from original series is stochastic series. Any modelling will now require the modelling of stochastic component. The modelling of stochastic component depends upon its statistical properties, probability distribution to which this series belongs and structure of the internal dependence (short term as well as long term). Whether a series is dependent or not is checked by correlation analysis, spectral analysis, run analysis and range analysis etc. These have been explained in detail by Lawrance and Kottegoda (1977) and Kumar (1982) and many other authors in different text books. An approach for time series modelling is described in the subsequent section.

1.2 Approach for Time Series Modelling

A systematic approach to hydrologic time series modelling is composed of following six main steps (Salas and Smith, 1980):

(i) identification of model composition,

(ii) selection of model type,

(iii) identification of model form,

(iv) estimation of model parameters,

(v) testing goodness of fit of the model and

(vi) evaluation of uncertainties.

Each of these are being described below:

1.2.1 Identification of model composition

In any modelling of hydrologic time series, one has to decide whether the model will be univariate or multivariate model or a combination of a multivariate and a disaggregation models etc. This decision is known as the identification of the model composition. Such identification generally depends on the characteristics of the overall water resources system, the characteristics of the hydrologic time series and the modeler's input. For instance, to analyse the operation of a reservoir by simulation, monthly inflows to such reservoir must be generated. If there is no other upstream reservoir or structures that may affect the operation of such reservoir, the univariate modelling of monthly streamflows at or near the site of the dam should be selected. On the other hand if other reservoirs exist or are planned upstream from the reservoir under study the multivariate modelling of monthly streamflows at various sites should be the choice. However instead of multivariate modelling of monthly streamflows, the modeller may select the multivariate modelling of annual series and then use disaggregation model to obtain the corresponding monthly flows. The above decision are contingent on the availability of adequate data in the system under study, as well as on their statistical characteristics. For instance, two time series which show cross correlation will require a bivariate modelling but if the cross correlation is not significant the two time series can be modelled independently by univariate models.

1.2.2 Selection of model type

Once the model composition is identified, the type of the model(s) must be selected. Namely the modeler has to decide on one among the various alternative models say AR (autoregressive), ARMA, ARIMA, FGN, BL (Broken line), SL (shifting level)or any other model that is available in stochastic hydrology. In this decision, these factors are important : the characteristics of hydrologic physical processes, the characteristics of hydrologic time series and the modeler's input.

1.2.3 Identification of model form

Once the type of model is selected the third phase of the modelling is to identify the form of the model. This identification as implied herein goes beyond determining the orders p and q say of an ARMA model as in the Box-Jenkins approach. For instance, in time series analysis of weekly streamflows, it is necessary to identify whether the series is skewed and if such skewness is constant or periodic whether the week to week correlation coefficients are periodic and whether the periodic characteristics should be described by the Fourier series, in addition to identifying the order say of an ARMA model.

1.2.4 Estimation of model parameters

Once the model is identified, the estimation of the parameters of the model is made. The method of moments and the (approximate) method of maximum likelihood are the two methods usually available. Generally the latter method gives the best estimates.

1.2.5 Goodness of fit of the model

The model estimated in phase (4) needs to be checked in order to verify whether it complies with certain assumptions about the model and to verify how well it represents the historical hydrologic time series. The model assumptions to be checked are usually the independence and normality of residuals of the model. In addition comparisons based on correlograms can be made to see if the model correlogram resembles the historical correlogram. Further comparison based on data generation, can be made to verify whether the model reproduces statistically historical statistics such as the means, variances, skewness, correlations, storage related statistaics, drought related statistics etc.

1.2.6 Evaluation of uncertainties

Once the model is judged to be adequate, it remains to evaluate the corresponding uncertainties i.e. (i) model uncertainty (ii) parameter uncertainty. Model uncertainty results because the true models of hydrologic time series are not known and at best the identified model composition, and selected type and form of the model are only close approximations.

Parameter uncertainty results because the model paramters are estimated from a limited amount of data . Model uncertainty may be evaluated by testing whether significant differences in the statistics generated by alternative models exist or not. Parameters uncertainty may be determined by finding the distribution of paramter estimates and by using the models with paramters sampled from such distributions.

1.3 Applicability of Time Series Models

Time series models have mainly two applications in hydrology and water resources i.e.(i) generation of synthetic time series and (ii) for forecasting future hydrologic series. Synthetic streamflows are required for reservoir sizing, risk analysis of water supply for irrigation systems, determination of rsik of failure of dependable capacities of hydroelectric projects and similar applications. Forecasting of hydrologic series are needed for short term planning of reservoir operation, real time and short term operation of river basins or systems for planning operation during an ongoing drought and similar operations.

2.0 REVIEW OF LITERATURE

2.1 Time Series Models in Hydrology

Development in the field of time series models started with the work of Hazen (1914) and Sudler (1927) who showed the feasibility of using statistics and probability theory theory in analysing river flow sequences. Hurst (1951) reported studies of long records of river flows and other geophysical time series. Barnes (1954) introduced the idea of synthetic generation of streamflow by using a table of random numbers.

Stochastic models proposed in the literature include: autoregressive (AR) models (Thomas and Fiering, 1962; Yevjevich, 1963; Matalas, 1967), periodic autoregressive (PAR) models (Jones and Brelsford, 1967; Pagano, 1978), contemporaneous autoregressive moving average (CARMA) models (Salas et al 1979), fractional Gaussian noise (FGN) models (Mandelbrot and Wallis, 1968; Matalas and Wallis, 1971), fast fractional Gaussian noise (FFGN) model (Mandelbrot, 1971), filtered fractional Gaussian noise model (Matalas, 1977), autoregressive moving average (ARMA) model (Carlson, et al, 1979; O'Connel, 1971), autoregressive integrated moving average (ARIMA) model, broken line (BL) model (Mejia, 1971), shot noise model (Weiss, 1973), model of intermittent process (Yakowitz, 1973; Kelman, 1977), disaggregation models (Valencia and Schake, 1973), Markov mixture models (Jackson, 1975), ARMA-Markov models (Lettenmaier and Burges, 1977), space time autoregressive moving average (STARMA) model (Deutsch and Ramos, 1984, Pfeifer and Deutsch, 1980), and general mixture models (Boes and Sals, 1978).

Although each model has its own merit and some of them can be successfully applied in hydrology yet they do have limitations. All of these have been criticized for one or more of the following reasons:

(i) not being able to reproduce short term dependence,

(ii) not being able to reproduce long term dependence,

(iii) difficulty in estimating the parameters,

(iv) limitations for generating large samples of synthetic data,

(v) lack of physical basis and

(vi) too many parameters.

2.2 Classification of Time Series Models

The first step in model construction is to select suitable classes or families of models from which the most appropriate model to fit to a given time series can be chosen by following the identification, estimation and diagnostic check stages of model development. For example when modelling annual hydrological time series one may wish to consider the ARMA family of models (Box and Jenkins, 1976), the classes of non Gaussian models suggested by Lewis (1985) and fractional differencing models (Hosking, 1985). Certainly if one is not aware that certain classes of models exist, one may not fit the most appropriate model to a given time series.

The time series models can be classified according to number of criteria. Hipel (1985) gives the following criteria for the classification of time series models:

(i) physically based and black box models

(ii) discrete and continuous time models,

(iii) continuous and discrete observations models

(iv) Gaussian and non-Gaussian models

(v) linear and nonlinear models

(vi) nonseasonal and seasonal models

(vii) stationary and nonstationary models

(viii) long and short memory models

(ix) multivariate and univariate models

(x) disaggregation and aggregation models

(xi) time domain and frequency domain models

(xii) Bayesian models

(xiii) state space formulation and Kalman filter.

The above criteria are explained in subsequent sections.

2.2.1 Physically based and black box models

Physically based models are designed to mathematically simulate the physical processes involved in the hydrological cycle. A physically based model can be deterministic or sometimes stochastic or it may contain both deterministic and stochastic components. Black box models don't consider the physical processes as such and are concerned with input and output only.

The problem with physically based model that it contain large number of parameters (Tong et al, 1985). Further due to great complexity of natural systems, the conceptual or physically based model is only a crude approximation to reality. Thompstone et al (1985) demonstrate that a simple stochastic TFN model forecasts more accurately than a cumbersome conceptual model which is very expensive to maintain and calibrate.

Even though most stochastic models were not originally designed to reflect the behaviour of physical phenomena a physical basis to these models can often be justified. For example, Salas and Smith (1981) demonstrate that a particular conceptual model of a watershed leads to ARMA streamflows and ARMA groundwater storage. Further discussions regarding physically based models are given by Klemes (1978) Yevjevich and Harmancioglu (1985) stress the importance of linking stochastic models with physically consistant properties of any particular water resources time series.

2.2.2 Discrete and continuous time models

The time variable in a stochastic model can be designed to handle discrete or continuous time. Most practical time series models are built for use with observations available at discrete time points. So most of the models are discrete time models.

2.2.3 Continuous and discrete observations models

Variables which are being observed such as riverflows and precipitation can be recorded as continuous or discrete variables. Most of the stochastic models have been developed for modelling a continuous variable available at discrete, evenly spaced times. Lewis (1985), Salas et al (1985), and Vecchia (1985) present a range of models for use with continuous variate time series. McKenzie (1985) describes models for which the variable being modelled is discrete.

2.2.4 Gaussian and non-Gaussian variabale models

Whether a model is Gaussian or not depends upon the variable. If the variable has been assumed to follow normal distribution than model is Gaussian variable model otherwise non-Gaussian.

Simple linear models such as the family of ARMA models are not necessarily defined as having Gaussian variates but are simplest to use as such because linear operations on Gaussian variates preserve Gaussianity or normality. Furthermore model construction procedures, based on the assumptions of Gaussianity are well developed. As a result, theoretical research regarding the development of stochastic models which can explicitly handle variables which are non-Gaussian and therefore don't follow a normal distribution has only been initiated recently. When the data are nonGaussian, one approach for obtaining data which are approximately normally distributed is to transform the original data, using a transformation such as a Box-Cox transformation (Box and Cox,1964) to produce a transformed series which is Gaussian. A model based upon the Gaussian assumption can then be fitted to the transformed series. An alternative approach is not to assume Gaussianity initially but to select a distribution that the original data actually follow. Lewis (1985) describe a range of new models developed for use with continuous variate nonGaussian time series. The distributions considered are Exponential,Gamma, Weibull,Laplace, Beta and mixed exponential distributions.

2.2.5 Linear and nonlinear models

When a model is linear, it is a linear function of the variables in the model. In a nonlinear model there is atleast one term where the variables and or innovations appear as products or are raised to powers. For instance, average daily riverflow data may have to be modelled using a model containing nonlinear terms because of the nonlinear relationship between runoff and precipitation over a small time scale. On the other hand a linear model may be sufficient to model mean annual river flows.

2.2.6 Nonseasonal and seasonal models

Nonseasonal models are designed for modelling time series which don't contain periodic components created by phenomena which are seasonal in nature. Certain types of geophysical records are strictly nonseasonal while in other situations it may be required to consider a time series of average annual values even if seasonal data are available. For example, tree ring indices and mud varve thickness are usually obtainable only in the form of yearly records, whereas mean annual riverflow, temperature and precipitation data can be calcualted from average weekly records.

Nonseasonal models can be fitted toyearly reocrds for use in various types of applications. For instance, when studying changes or trends in the climate over a long time span it may be advantageous to analyse annual time series. By using transfer function noise models (TFN) to link annual riverflows, precipitation and temperature with tree ring data for intervals at which the time series overlap, the calibrated TFN model can be used to back forecast the missing observations in the hydrological time series.

Seasonality consists of different kinds of periodic behaviour in a given phenomenon which are caused by natural or man induced events. For instance, due to the annual rotation of the earth about the sun and the concurrent changes in tilting of the earth's axis, weather characteristics at a given location on the earth usualy possess pronounced seasonal characteristics. The seasonal properties of precipitation and temperature, in turn, impart periodic behaviour on the entire hydrologic cycle. Riverflows in the Indian subcontinent, for example, are much higher during monsoon rains. Besides natural causes for seasonality in time series , man's activities can create significant seasonal behaviour in time series. For example, agricultural, residential and industrial water demands are often dependant upon the time of the year. Because a man induced seasonaility may not occur at precisely the same time every year, it is sometimes necessary to model the starting times of various kinds of seasonalities using random variables.

Yevjevich and Harmancioglu (1985) and Salas et al (1980) discuss the different kinds of periodic behaviour of hydrological systems and general ways the measurements from these natural systems can be appropriately modelled. Periodic autoregressive moving average (PARMA) and periodic autoregressive (PAR) models are specifically designed for modelling seasonal hydrological time series. Most of the nonseasonal models can be extended for use with seasonal data by deseasonalizing the data. To economize on the number of parameters required in the deseasonalization procedure, a Fourier series approach can be employed (Rao et al, 1985; Tao and Delleur, 1976; Salas et al , 1980; McKenzie,1985) whatever is the case, subsequent to deseasonalization a model such as a nonseasonal ARMA or a nonseasonal fractional differencing model (Hosking, 1985) can be fitted to the deseasonalized data.

2.2.7 Stationary and nonstationary models

Stationarity of a stochastic process can be qualitatively interpreted as a form of statistical equilibrium. Therefore, the statistical properties of the process are not a function of time. Besides reducing the mathematical complexity of a stochastic model, the stationarity assumption may reflect reality. For instance, if a natural river basin has not been subjected to any land use changes such as urbanization and cultivation, it may be reasonable to assume that a stationary stochastic model can be fitted to the time series of historical annual riverflows.

In certain situations, the statistical characteristics

of a p

of a process are a function of time. Water demand tends to increase over the years as metropolitan areas grow in size and the affluence of the the individual citizen increases. The average carbondioxide content of the atmosphere may increase with time due to complex natural processes and industrial activities. To model an observed time series that possesses nonstationarity, a common procedure is to first remove the nonstationarity by invoking a suitable transformation and then to fit a stationary stochastic model to the transformed series. For instance , Box and Jenkins (1976) suggest differencing the given data to remove homogeneous nonstationarity before designing an appropriate stationary model such as an ARMA model. Therefore, even when modelling nonstationary data, the mathematical results that are available for describing stationary processes are often required.

Some researchers believe that natural processes are inherently nonstationary and therefore the greater the time span of the historical series, the greater is the probability that the series will exhibit statistical characteristics which change with time. However, for relatively short time spans it may be feasible to approximately model the given data sequence using a stationary stochastic model.

2.2.8 Long and short memory models

When considering a stationary time series represented by z_t at time t, the autocovariance function or equivalently the ACF can be employed to measure the linear dependence between observations. The covariance between Z_t and a value Z_{t+k} which is k time lags removed from Z_t is theoretically defined in terms of the auto covariance at lag K given by

$$\gamma_k = \mathrm{Cov}\ [Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] \qquad \ldots(2)$$

A normalized quantity that is more convenient to deal with than γ_k is the theoretical ACF which is defined at lag k as

$$\rho_k = \frac{\gamma_k}{\gamma_o}$$

The theoretical ACF(ρ_k) is dimensionless and therefore independent of the scale of measurement. Furthermore, the possible values of ρ_k range from -1 to +1 and ρ_k is unity at lag zero.

For a known stochastic process, such as an ARMA process, it is usually possible to calculate the theoretical ACF. An important measure of the appropriateness of a model, which has been fitted to a given time series is to see if the theorertical ACF of the fitted model statistically resembles the sample ACF of the data.

When the theoretical ACF is summable it must satisfy (Brillinger,1975)

$$\sum_{-\infty}^{\infty} |\rho_k| < \infty \qquad \ldots(3)$$

A covariance stationary process is said to possess a short or long memory according to whether or not the theoretical ACF is summable. Examples of short memory processes are the stationary ARMA processes of Box and Jenkins(1976) whereas the fractional Gaussian noise model (Mandelbrot and Wallis,1969) possesses long memory for a specified range of a model parameter. The importance of both long and short memory processes for modelling annual hydrological time series is exemplified by the study of the 'Hurst Phenomenon'. Hosking (1985) briefly describes the Hurst phenomeon and explains how the fractional differencing model can be used to model both long and short memory.

2.2.9 Multivariate and univariate models

Salas et al (1985) present an extensive survey of multivariate modelling of water resources time series. The term multivariate is employed because at least two time series are modelled when using a multivariate model. Qualitatively, a multivariate model can be written as

Multiple outputs = Dynamic component + Noise component

Where the dynamic component models the way in which all of the series dynamically influence one another and the noise component models the correlated portion of all the series which is not modelled by the dynamic component. Because the behaviour of each of the multiple series is dependent over time upon the other series, the overall multivariate model is often referred to as a dynamic model.

An univariate model is a model for which there is a single output but there may be one or more inputs plus a noise component. A special type of univariate model for describing a single time series is the ARMA model.

2.2.10 Disaggregation and aggregation models

Another special class of multivariate models is the disaggregation model proposed by Valencia and Schake (1973). This model allows one to break down a series for which there are longer time units separating values into a sequence of values separated by shorte time units. For instance, an annual series can be disaggregated into a monthly series. In disaggregation relevant statistical properties remain consistent at both the annual and seasonal level. Annual flows can be generated by a short or long memory model and these annual flows can then be disaggregated to the seasonal flows.

Many researchers believe that there is not much information contained in the annual series, a better procedure may be to aggregate rather than to disaggregate. Vecchia et al (1983) present a convincing argument which favours the concept of aggregation over disaggregation. They prove that, if the original seasonal data follow a PARMA model with one moving average and one autoregressive parameter (i.e., PARMA(1,1)), then the aggregated annual data must be an ARMA model with one AR and one moving average parameter (i.e., ARMA(1,1)). Furthermore there is significant gain in parameter estimation efficiency at the aggregated level when the seasonal data and their model is used rather than the aggregated (i.e.annual) data and their model.

2.2.11 Time and frequency domain models

In order to fit a time series model to a data set various techniques are available for use at the three stages i.e. identification, estimation and diagnostic check stages of model construction. If a given method or statistic such as the sample ACF which can be used for model identification is expressed directly in terms of the time variable one is said to be working in the time domain. Alternatively, one can work in the frequency domain also by entertaining Fourier transforms. The Fourier transform of the autocovariance function produces the spectrum which expresses the distribution of the variance of the series with frequency (Jenkins and Watts, 1968). Brillinger (1985) presents interesting results regarding Fourier inference.

2.2.12 Bayesian models

By employing Bayesian approaches, one can take various types of prior information into account in a time series study. Kryzysztofowicz (1985) explains how one can obtain optimal forecasts of hydrologic time series by employing the Bayesian processors of forecasts.

2.2.13 State-space formulation and the Kalman filter

Another general set of concepts which can be utilized with most time series models is the state space formulation of models and the Kalman filter. As explained by Bergman and Delleur (1985a) a given model can be transformed into a state-space formulation for which the state variable vector can be continuously caliberated as new data become available by using the discrete Kalman filter algorithm. Bergman and Delleur (1985b) apply state space techniques and Kalman filtering to the class of autoregressive models for the purpose of real time flow forecasting. They present an adaptive Kalman filter algorithm which is evaluated using simulation. They apply it for the daily streamflow forecasting of the Potmac River.

As is clear from the above discussion,discrete stochastic models may be classified in may ways. Here it has been subjectively decided to classify them as short memory or long memory models. Attempts to reproduce long term characteristics such as the Hurst phenomena distinguish the long memory from short memory models. Short memory models of hydrologic phenomena include the moving average (MA) models, autoregressive (AR) models and mixed autoregressive moving average (ARMA) models. Each of these models is sequential and may be univariate or multivariate. Long memory models include fractional Gaussian-noise models, filtered fractional Gaussian noise models, broken line models, and depending upon parameter values, certain autoregressive integrated moving average (ARIMA) models. They may also be univariate or multivariate and are sequential. A non-sequential model that may have either long or short memory and be univariate or multivariate is the disaggregation model. All of these models are reviewed and described briefly in forthcoming sections. Daily flow generating model and multisite data generating models have also been explained in the end.

2.3 Short Memory Models

The use of short memory models in hydrologic analysis was introduced primarily to produce synthetic sequences of flows to route through a water resources system, the idea being to test it under 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 (1967), Hufschmidt and Fiering (1966), Schaake and Fiering (1967) Davis(1968), Hall et al (1969), Morean and Pyatt (1970), Hamlin and Kottegoda (1971,73), Gupta and Fordham (1972), Hamlin et al (1973,75) , Spolia and Chander (1974) and many others. Most of t se 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 Lenton and Rodriguez Iturbe, 1974 and Klemes, 1979.

Short memory models have also extensively been used for forecasting flows. If the series is nonseasonal 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 models 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), Clarke (1973), Delleur and Kavass (1978), Chander et al (1980). Recently the use of control engineering concepts have been introduced in the 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 Egly, 1977), flood evaluation (Kottegoda, 1972, 73) etc.

Commonly used short memory models to hydrologic Time series, modelling are described in following sections.

2.3.1 Moving average model

The moving average model is the simplest short memory model and expresses a sequence of events (i.e. annual flows) in terms of deviations at time t, $\tilde{Z}_t$, from the mean, μ of the process or sequence of events, Z_t .

$$\tilde{Z}_t = Z_t - \mu \qquad \text{... (4)}$$

The deviation from the mean of the process is expressed as a finite weighted sum of elements plus a random element a_t.

$$Z_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad \cdots\cdots - \theta_q a_{t-q} \qquad \ldots(5)$$

in which the θ_i are the weighted parameters and the a_{t-i} are random elements (white noise). The above equation represents a moving average process of order q. It embodies q+2 parameters, μ, $\theta_1 \ldots\ldots, \theta_q$ and the variance of a_i, σa^2, that must be estimated from data sets, before it can be used in practice.

2.3.2 Autoregressive model

Another model that has been used rather extensively in hydrologic analysis is the autoregressive model. It is a very useful tool in the simulation of hydrological and climatological data. As with the moving average process, this model also works with deviation, $\tilde{Z}_t$ from the mean, μ of the process or sequence of events, Z_t. However, the autoregressive process expresses the deviation from the mean of the process, a finite weighted sum of previous deviation plus a random variate, a_t. Thus

$$Z_t - \mu = a_t + \theta_1 (Z_{t-1} - \mu) + \theta_2 (Z_{t-2} - \mu) + \ldots\ldots \quad + \theta_p (Z_{t-p} - \mu) \quad \ldots (6)$$

is an autoregressive process of order p. It contains p+2 parameters, μ, $\theta_1 \ldots\ldots \theta_p$ and σ_a^2 that must be estimated from a given data. The mean of the sequence of events is μ, the weight factors are θ_i and the variance of random variates is σ_a^2.

Matalas (1977) discuss the autoregressive model and some of its weaknesses as well as corrections that should be applied to the coefficients to correct for bias in estimates of moments.

2.3.3 Mixed autoregressive moving average model

In practical solution to many hydrologic problems it may be necessary to include both autoregressive and moving average terms to obtain a parsimonious model. Box and Jenkins (1976) describe such a composite model by expressing the deviation of a variate from its mean, Z_t, as a finite weighted sum of previous deviations plus a finite weighted sum of random variates plus a random element a_t. Thus:

$$\tilde{Z}_t - \theta_1 \tilde{Z}_{t-1} \cdots\cdots\cdots - \theta_p \tilde{Z}_{t-p} = a_t - \theta_1 a_{t-1} \cdots\cdots - \theta_q a_{t-q}$$

is an autoregressive moving average (ARMA(p,q)) model of autoregressive order p and moving average order q.

2.3.4 Autoregressive integrated moving average model

Hydrologic data such as flow records, temperature, rainfall etc. exhibit seasonal and other cyclic patterns. These patterns can be eliminated from the data by seasonal transformation to zero mean and unit variance:

$$Z_t = \frac{q_t - q_j}{\sigma_j}$$

where j denotes the j^{th} period in the cycle. Such transformations require many parameters especially if monthly patterns are present.

Both the effects of a trend that is not seasonal and cyclic seasonal patterns in the mean can be removed by a proper differencing of the data set, i.e. the substraction of a value from its previous value or its value j units apart. If this has been done properly, then the data set will be a stationery series and the moving average, autoregressive or the mixed autoregressive moving average models can be used to describe the process. In this case, mixed model is known as an autoregressive integrated moving average (ARIMA) model.

Box and Jenkins (1976) discuss the ARIMA (p,d,q) model expressing the d^{th} order differences of a series Z_t as ARMA(p,q) model. In general, this type of model is of much more value in forecasting than in simulation. O'Connel (1977) discusses application of ARIMA type models in synthetic hydrology. He states that in general most hydrologic phenomena, particularly runoff can be expressed as an ARIMA (1,0,1) or ARMA (1,1) process. This process should resemble the historic sequence in terms of the mean, variance, lag one autocorrelation and Hurst phenomenon.

Some researchers have also looked at use of ARIMA models in forecasting. Mckerchar and Delleur (1974) compares a second order autoregressive model with 27 parameters, operating on standardized monthly flow data to an ARIMA model that required only 4 parameters. Forecasting with both models tended toward the monthly means as the lead time increased. The autoregressive model forecasts also tended toward the monthly standard deviations. However, because the ARIMA model did not account for seasonal variability in monthly standard deviations, the forecasting errors could not be associated with physical reality.

Mejia et al (1975) demonstrated the use of ARIMA type models in simulating or predicting fluctuations in water quality parameters of the Passaic river. Different forms of the ARIMA model were best suited to different parameters. Daily flow residuals were represented by an ARIMA (2,1,0) model, daily water temperature by an ARIMA (1,0,1) model, and daily Biochemical oxygen demand and oxygen deficit by ARIMA(1,0,0) models.

2.4 Long Memory Models

Long memory odels are specifically designed to reproduce the Hurst Phenomenon (Mandelbrot and wallis,1968). This is accomplished by several means even though the concepts or causes of the phenomenon are not known. The Markov models or autoregressive models can faithfully produce the high frequency components of the data sets but fail to produce the extremely low frequency components typified by the Hurst phenomenon Multilag models have attempted to account for the low frequency components but they are generally not satisfactory for the very low frequency response.

Discrete Fractional Gaussian Noise (DFGN) is one of the approaches considered. It is Gaussian random processes with a k^{th} order autocorrelation coefficient (Mandelbrot,1971) given by

$$\rho(k) = \frac{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}}{2}$$

It has a single parameter, H, the Hurst coefficient, which in most hydrologic applications has a value in the interval $0.5 \leq H \leq 1.0$. The above process was invented by Mandelbrot to model the variance of phenomena characterized by long run effects in which the cumulative influence of very small serial correlations between remote values is non-negligible.

Construction of a sample function of DFGN involves the summation of an infinite number of components. Therefore, approximations of DFGN are used to limit the number of components. However, the number of operations must be high enough to preserve the desired value of H if long term properties of a series are of interest. The order required tends to increase in proportion to the total length of period

T, to be simulated.

Fast fraction Gaussian Noise (FFGN) processes (Mandelbrot, 1971), Broken line processes (Mejia et al,1972) and ARIMA processes are used to approximate DFGN.

2.4.1 Fast fractional Gaussian noise processes

Fast fractional Gaussian noise (FFGN) models are discrete approximation to theoretical fractional Gaussian noise. They are made up of three components:(i) An independent autoregressive processes, $\rho_h X_h(t-1)$, used to obtain the high frequency effects, not present in the low frequency term, but necessary for discrete time fractional Brownian motion; (ii) A low frequency or long-run-effects term to reproduce the low frequency properties of the covariance function. formed by superimposing weighted outputs from a parallel set of first order autoregressive processes, $\Sigma W_j(\rho_j x_j(t-1)+\varepsilon_j(t))$ and (iii) A random element $\varepsilon_h(t)$

The model is defined as

$$X(H,t)= \rho_h X_h(t-1) + \sum_{j=1}^{N} W_j(\rho_j X_j(t-1)+ \varepsilon_j(t))+ \varepsilon_h(t) \qquad \cdots \qquad (10)$$

Here X(H,t) has mean 0 and variance 1; ε_h and ε_j are independent Gaussian processes of mean 0 and variance 1, ρ_h is a function of the Hurst coefficient, H, the number of terms used in the approximation, N, and a 'base' value, β; ρ_j is a function of β only; and W_j is the weighting factor for the j^{th} autoregrssive process Each of the autoregressive processes in the long run effects term have successively longer memories.

The number of AR(1) processes required for the long run effects term is expressed as;

$$N = \log(QT)/\log(\beta) \qquad \ldots (11)$$

Where T is the maximum duration of the period of interest, β is the base value ($\beta > 1$) controlling the separation of individual AR(1) 'decay' parameters, and Q is a 'quality' parameter controlling the number of AR(1) processes needed for a given β value. As β approaches 1, Q increases and the approximation becomes more accurate because more AR(1) processes are included. Sufficient accuracy for most practical applications should be achieved if Q = 6 and β = 3. If these values are used, then

$$N \simeq 2 \log (6T)$$

The weights applied to each autoregressive subsystem output are obtained from

$$W_n^2 = \frac{H(2H-1)\ (\beta^{1-H} - \beta^{H-1})}{\Gamma(3-2H)}\ \beta^{2(H-1)n} \quad ; 1 \leq n \leq N \qquad \ldots (12)$$

where Γ denotes the Gamma function.

The autoregressive correlation parameter is

$$\rho_n = \exp(-\beta^{-n}) \qquad \ldots. (13)$$

Chi et al (1973) demonostrate how to use the model in simulation and describe all of the steps necessary. Using these steps, the FFGN is probably the best of the discrete fractional Gaussian noise models.

2.4.2 Filtered fractional noise processes

The sequence of values, $X_1, X_2 \ldots\ldots$ of a filtered fractional noise process, are generated by applying a set of weights, W_i, successively to a sequence of independent random variable, $\varepsilon(i)$. The generating equation for the value of X at time t is

$$X_t = (H-0.5) \sum_{i=pt-M}^{pt-1} W_i \, \varepsilon(i) \quad ; \quad 0.5 < h < 1.0 \qquad \ldots(14)$$

where,

$$W_i = (pt-i)^{H-1.5}$$

where M, the memory of the system, (usually much longer than the period to be generated, that is several thousand) is a function of Hurst coefficient, H, and the lag-K autocorrelation coefficient, p is an integer greater than one. The weights, $(pt-i)^{H-1.5}$, vary in value from nearly zero to one. Matalas and Wallis (1971) derived values of the mean, variance, skewness and lag one autocorrelation coefficient for this model and expressed equation (14) in terms of these paramters. Estimated values of parameters plus the value of H may then be used to generate synthetic sequences. However if short sequences are generated, values of variance, skew, lag one autocorrelation and Hurst coefficient must be adjusted for bias as described by Matalas (1977). A major disadvantage of using this model is that it is computationally expensive because of the large number of terms added to produce the synthetic sequence. H may take on values from 1000 to as high as 50,000.

2.4.3 Broken line processes:

A broken line process consists of the summation of a finite number of simple broken line processes (Mejia et al,1972). The simple broken line process (Fig.1) is a sequence of intersecting line segments in which the time projections between intersections are of same length a. The values of the process $\varepsilon(t)$ at the intersections are independent and are identically distributed random variables, π with zero mean and unit variance, A simple broken line process is given by

$$\beta_i(t) = \{\eta_m + \frac{\eta_{m+1} - \eta_m}{a} (t-ma)\ I\} \qquad \ldots (15)$$

where,

η_m = independently identically distributed random number of zero mean and variance σ^2,

a = time distance among η_m

I = 1 when ma < t < (m+1) a

0 otherwise

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

$$\rho_k = 1-0.75\ (k/a)^2\ (2-(k/a)) \quad \text{for } 0 < k < a \qquad \ldots (16)$$

$$\rho_k = 0.25\ (2-k/a)^3 \quad \text{for } k > a\ ; \qquad \ldots (17)$$

$$\rho_k = 0 \quad \text{for } a < k/2 \qquad \ldots (18)$$

for modelling, a BL process is formed by adding a finite number broken line $\beta_i(t)$

$$a_t = \sum_{i=1}^{n} \beta_i(t) \qquad \ldots (19)$$

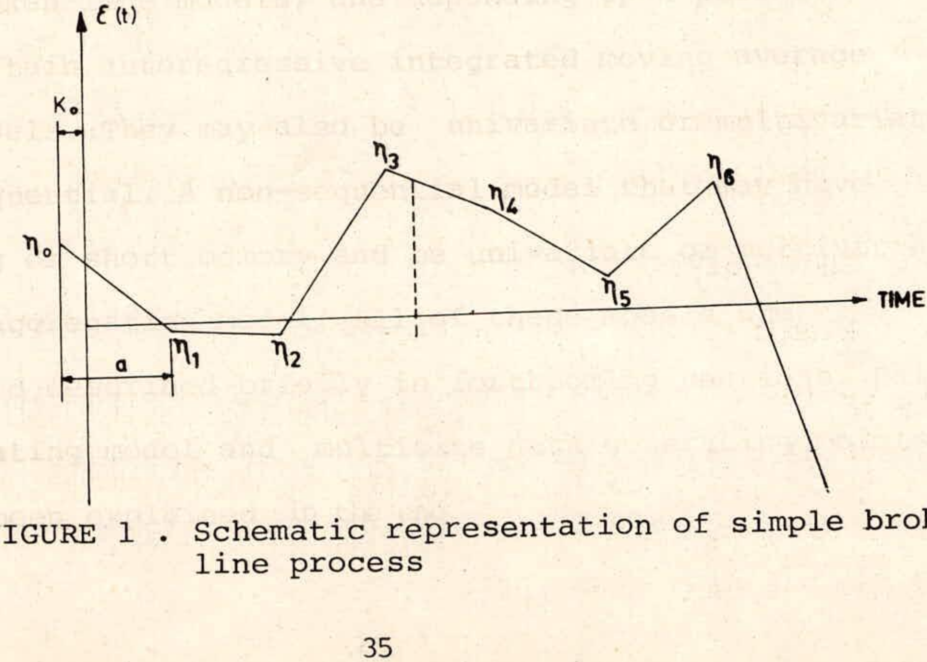

FIGURE 1 . Schematic representation of simple broken line process

Mejia (1972) has related the parameters η ,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 to model it. FFGN models therefore been preferred compared to BL for long memory modelling.

2.5 Daily flow generating models

The modelling of realistic daily flow sequences is slightly complex because of three reasons ,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 there is no flow. 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 cannot produce sharp rising limbs and slow recession Kumar(1982).

Weiss(1973a,b) suggested a daily flow generating model known as ' Shot noise model' which has a built in capacity to model the ascension recession behaviour. The model is briefly explained in the following section:

2.5.1 Shot noise Model

Examination of a continuous streamflow record shows series of spikes of various heights, followed by exponential decays. Weiss (1973a) used this observation in proposing the use of shot noise model to represent daily flow records as a stochastic process.

The shot noise model is a continuous model that assumes the temporal distribution between events and the magnitude of the jump inflow values associated with each event, the peak or spike, are exponentially distributed(Fig.2a to2d) taken from O'Connell,1977). The absolute value of the flow on any given data is assumed to be fraction of all previous events(the flow rate on the previous day multiplied by a decay parameter) plus any increase from an event that may have occurred since the previous day. Since the process is continuous in time more than one event may be generated in the time interval of one day. The generating equations accumulate these and then discrete values, which are assumed to be based on a daily average, are calculated.

To use the model in synthetic data generation, two highly interrelated processes must be modelled. One is the incremental increase in flow rate denoted by x_{t+1}, and the other is the discretized value of the continuous process denoted by, x(t+1), for example, the daily flows. The numerical increase in flow rate is given by

$$x_{t+1} = 1/b \ (1-e^{-b})\, x(t) + \sum_{m=N(t)}^{N(t+1)} (1/b)\, Y_m \ (1-e^{(-b(t+1-\tau_m))}) \qquad ....(20)$$

and the discretized continuous (daily) process is given by

$$x(t+1) = e^{-b}\, x(t) + \sum_{m=N(t)}^{N(t+1)} e^{(-b(t+1-\tau_m))}\, Y_m \qquad ....(21)$$

In equations 20 and 21

b is the decay rate;

τ_m is the time of an event within the time interval of 1 day and is calculated as $\tau_{m+1} = \tau_m + E$;

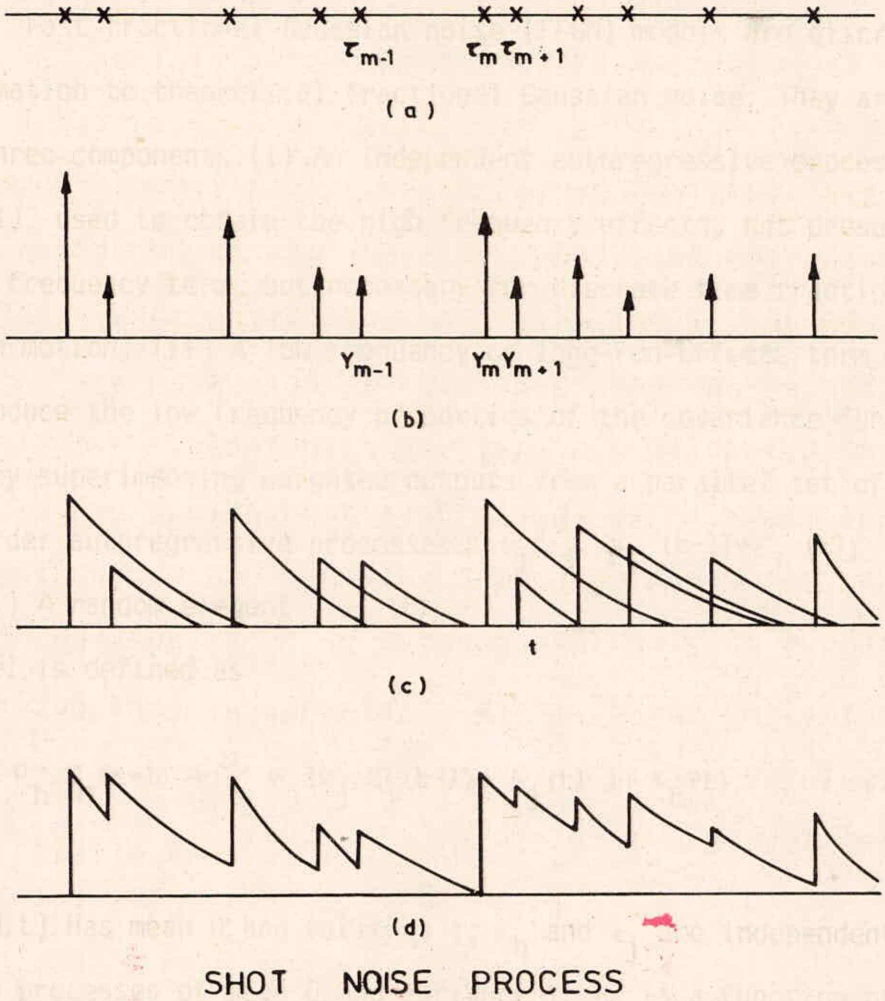

Fig.2 -Shot Noise Process (a) Events..., τ_m..... from a Poisson process with rate ν . (b) Jumps ...., Y_m .... from an exponential distribution with mean θ. (c) Pulses with values ....,

$Y_m e^{-b(t-\tau_m)}$ ...., at time t. (d) Schematic plot of continuous single shot noise process (From O'Connell, 1977, Fig.2).

E is pseudorandomly generated from an exponential distribution with mean $1/\upsilon$;

Y_m is the magnitude of the spike at time τ_m and is generated from an exponential distribution with mean θ ;

m is a number used to keep track of the number of events, if any, within the daily interval t to t+1;

N(t) is a Poisson process with rate υ .

The first term in equation 20 and 21 is the effect of previous events and the second is the innovation term. Both the equations describe a continuous time process. Daily flow values which are used in analysis are average values of 24 values. Values of υ,θ and b are calculated from moments of mean daily flows as

$$\mu_x = \frac{\upsilon\theta}{b} \qquad \text{...(22)}$$

$$\sigma_x{}^2 = \frac{\upsilon\theta}{b}\left(\frac{2(b-(1-e^{-b}))}{b^2}\right) \qquad \text{....(23)}$$

and

$$\rho_x(1) = \frac{(1-e^{-b})^2}{2(b-(1-e^{-b}))} \qquad \text{....(24)}$$

In equations 22,23,and 24 μ_x , $\sigma_x{}^2$ and $\rho_x(1)$ are the mean ,variance and lag 1 autocorrelation coefficients respectively of the mean daily flow values (Weiss, 1973b; O'Connell,1977).

In dealing with instantaneous flows equation 22 to 24 should not be used. Modified equations given by Weiss(1977) should be used.

Weiss (1977) also suggested a double shot noise model, one to represent direct runoff for rainfall & other one to account for ground water storage. If used in seasonal flow modelling obviously, the value of and b may vary from month to month. Fitting of such a model to daily hydrologic data is quite complex and is a laborious task. So one should consider the use of a watershed model with stochastic input rather than a completely stochastic model of daily flow.

2.6 Disaggregation Model

Disaggregation model is a general autoregressive model of which all other autoregressive models are a special case. This is capable of using the output of any model as input and generating a parallel series of seasonal events that aggregate to the original series. Furthermore the model may feed upon itself to further disaggregate the series to monthly, weekly daily and hourly events.Structure of the output series is based entirely on the statistical characteristics of the originaal data set; thus the model cannot reproduce physical characteristics of a daily streamflow series but it can be used for the generation of hourly rainfall.

Fundamentally the model disaggregates the annual series into seasonal values by using statistical data that show what fraction of the annual series is attributed to each season. It is made up of two components, a deterministic component that is proportional to the correlation of each season with the annual values and a random component.

The disaggregation model takes the form (Valencia and Schaake,1973).

$$Y(t) = A\ X(t) - BV\ (t) \qquad \ldots(25)$$

where,

$Y(t)$ is an n x 1 vector of seasonal values for the tth year.

n is the number of seasons in a year (for example 12 if monthly values are to be generated);

$X(t)$ is the value of the annual series for the tth year

$V(t)$ is an n x 1 vector of independently distributed standard normal deviates;

A is an n x 1 vector of coefficients

B is an n x n matrix of coefficients

The vector A and matrix B are obtained by analysis of N years of the historical data that relate the seasonal to annual values

$$\hat{A} = \hat{S}_{yx} \hat{S}_{xx}^{-1} \quad \text{....(26)}$$

where,

$\hat{A}$ is estimated A vector,

$\hat{S}_{yx}$ is the (nx1) cross products matrix(proportional to the covariance matrix) between n seasonal observation,Y and the annual values X summed over the N year of record used to evaluate the coefficients;

$\hat{S}_{xx}$ is the inverse of the sum of squares matrix of the N years of observed annual values,X.

$$\hat{B} \hat{B}^T = \hat{S}_{yy} - \hat{S}_{yx} \hat{S}_{xx}^{-1} \hat{S}_{xy} \quad \text{....(27)}$$

where,

$\hat{B} \hat{B}^T$ is the best estimate of the product of the B matrix and its transpose;

$\hat{S}_{yy}$ is the sum of squares matrix over N years of the n seasonal values Y;

$\hat{S}_{yx}$ and $\hat{S}_{xx}$ as described above

S_{xy} is the cross products matrix between the annual values X and n seasonal values Y summed over N years.
The matrix B can be calculated from $\hat{B}B^T$ by the technique of principle component analysis (Matals, 1967)
Valencia and Schaake (1973) show that the parameter matrix A and B preserve continuity between seasonal and annual values;

$$X_i = \sum_{j=1}^{n} Y_{ij} \qquad \text{...(28)}$$

Equations 26 and 27 are used with historic data to solve the A and B matrices respectively. These matrices are used in equation 25 alongwith a random number generator and the annual series to develop seasonal values for the annual series. The values may further be disaggregated to form sequences of any desired time interval.

Disaggregation model can be extended for multisite data generation also (Tao and Delleur, 1976). In this case the vectors X and Y become matrices of annual and seasonal, respectively, extending across as many stations as desired. Vector A becomes a matrix of coefficients that disaggregated each of the station's annual series into average seasonal values. Matrix B is a coefficient matrix that preserves the covariance and corss-covariance structure of the residuals introduced in the vector V. They used the model for both single and multistation disaggregation and evaluated the model performance with respect to preservation of both first and second order moments. Tao and Delleur (1976) also showed how the model can be used to generate annual and monthly runoff volumes from annual rainfall. Mejia and Roussele (1976) modified the disaggregation model to preserve correlations of the last season of one year with seasons of the following year.

The model can be used alongwith any annual data generation model. The model is specially useful where the long range properties of the annual series are of interest. A major disadvantage of the model is the need to calculate the large number of parameters for the A & B matrices.

2.7 Multisite models

Multisite models are used when joint modelling of hydrologic sequences of several sites is required. Sometimes it is not sufficient simply to model flows at every site independently because the flows at various sites are strongly interrelated. The simultaneous modelling of flows at more than one site at a time began with the key and sattelite approach of Thomas and Fiering (1962) which took account of cross correlations between the series but did not model the serial correlation at the satellite site. Later Fiering (1964) approached the general multisite model using a principle component analysis. The classic appraisal and development of Fiering's work by Matalas (1967) forms the basis of presently used multisite lag 1 autoregressive models, such as Young and Pisano (1968) and Moreau and Pyatt (1970). Lawrance (1976) improved upon the Fiering model in the sense that lag 0 and lag one cross correlation coefficients are maintained in the model. Multisite models can be classified as (i)Multisite short memory models and (ii) Multisite long memory models.

Multisite short memory models include multisite lag one and higher order autoregressive models. Multisite higher order autoresseive models are required if serial and cross correlation coefficients at lags more than one between sites are to be maintained. A good dicription of multisite short memory models is given by Kumar (1982), Lawrance and Kottegoda (1977).

In order to simplify the estimation procedure in multisite short memory models, decoupled multisite modesl 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 standardized 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. The residual series will be serially independent but will have cross correlation with residual series at other stations. 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.

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) for the fractional noise process , Mejia et al (1974) for broken line process, O'Connell (1974), for the ARMA models and Weiss (1977) for shot noise models. 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.

3.0 REMARKS

The literature available on the application of time series models is enormous. In this review note, time series models have been briefly reviewed. The main emphasis has been given on fundamental structure of various models and techniques for time series modelling have only been touched. Review of rainfall generating models has not been attempted. Some of the areas in which further work is required are as follows:-

1. There is need to develop more comprehensive families of models that can simultaneously handle a wider variety of the criteria given in section 2.2. For example it may be possible to design a multivariate model that can handle both non linear and non Gaussian characteristics of data. However, any new class of models should be designed to be as simple as possible and thereby not have too many parameters.
2. Efforts should be made to incorporate both the physical and statistical aspects of the problem into the basic model design.
3. 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 developement especially

in reducing the number of parameters to be calculated from the historical records. Application of shot noise model for Indian rivers will be highly informative.

4. Further work is required regarding use of a watershed model with stochastic input.

5. 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.

6. In depth study of short memory and long memory models is required. The drawback of short memory models is that they don't model Hurst h. Long memory models do model Hurtst h but it is only a mathematical excercise without any physical meaning.

7. There is need of work in water quality and quantity time series analysis to meet the ongoing concern for the solution of complex environmental problems.

8. The use of time series models for modelling monthly/pentad flows for Indian rivers may require some modifications as many Indian rivers have nearly zero flows during non-monsoon season (November-May). Case studies regarding application of time series models to Indian rivers will be highly informative and good contribution to hydrology literature as very little has been done in the application side.

REFERENCES

1. Bergman,M.J. and Delleur,J.W.(1985a),' Kalman filter estimation and prediction of daily streamflows,I. review,algorithm, and simulation experiments', Water resources bulletin,Vol.21/5, pp.815-826.

2. Bergman,M.J. and Delleur,J.W.(1985b),' Kalman filter estimation and prediction of daily streamflows II. Application to the Potmac River', Water Resources Bulletin,Vol.21/5,pp.827-832.

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

4. Boes, D.C. and Salas,J.D.(1978),' Nonstationarity in the the mean and the Hurst phenomenon', Water Resources Research,Vol.14/1,pp.135-143.

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

6. Box,G.E.P. and Jenkins,G.M.(1976),' Time series analysis, forecasting and control', Hoden Dey Publishing Co., California.

7. Brier, G.W.(1961),' Some statistical aspects of long term fluctuation in solar and atmospheric phenomenon', Annals New York Academic Sciences,Vol.95,pp 173-187.

8. Brillinger,D.R.(1975),' Time series: data analysis and theory', Holt Rinehart and Winston,Newyork,New york.

9. Brillinger,D.R.(1985),' Fourier inference: some methods for the analysis of array and nonGaussian series data,' Water Res.Bulletin,Vol.21/5,pp.743-756.

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

11. Buishand T.E.(1977),' Stochastic modelling of daily rainfall sequences',Meddleingen Landbouewhogeschcool Wageningen,Netherlands,Vol.77.

12. Carlson,R.E., Mac Cormick,A.J.A. and Watts,G.G.(1970), " Application of Linear random models to four annual

stream flow series', Water Res.Res.,Vol.6,pp.1070-78.

13. Chander,S ., Spolia,S.K., and Kumar,A(1980),' Autorun analysis of hydrologic time series- A comment,' Jnl. of Hydrology,Vol.45,pp.33- 39.

14. 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.

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

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

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

18. 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.

19. Deutsch,S.J. and Pfeifer,P.E.(1981),' Space-Time modelling with contemp-oraneous correlated Innovations', Technometrics,Vol.22.

20. Deutsch,S.J. and Ramos,J.A.(1984),' Space time evaluation of reservoir regulation policies', Technical Report,School of Civil Engineering,Georgia Institute of Technology.

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

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

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

24. 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.

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

26. Hamlin,M.J. and Kottegoda,N.T.(1973),' The preparation of a data set for hydrological system analysis', Design of Water Resources Projects with Inadequate Data, Symposium, Madrid (IAHS,Publ.No.108,1974)Vol.1 pp.305-313.

27. Hamlin,M.J.,Fisher,R., and Cluckie,I.D.(1975), ' Multisite data generation for large water resources systems', IAHS symposijm of Bratislava,Application of Mathematical Models in Hydrology and Water Res. Systems, Workshop paper,.

28. Hamlin,M.J., Kottgoda, N.T. and Kitson,T.(1973), Control of a river system with two reservoirs', (In symposium on control of Water Resources Systems Int. Fed. of Automat.Contr.,Haifa,Israel) Journal of Hydrology,(1976), Vol.28,pp.155-173.

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

30. Hazen,A.(1914),' Storage to be provided in impounding reservoirs for municipal water supply', Trans. Amer. Soc.Civil Engineers,Vol.77,pp.1539-1669.

31. Hipel,K.W.(1985),' Time series analysis in perspective Water Resources Bulletin,Vol.21/4,pp.609-624.

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

33. Hosking,J.R.M.(1985),' Fractional differencing modelling in hydrology,Water Resources Bulletin Vol.21/4,pp.677-682.

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

35. Jackson,B.B.(1975),' Markov mixture models for drought lengths', Water Resources Research,Vol. 11/1,pp.64-74.

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

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

38. Kelman,J.(1977),' Stochastic modelling of hydrologic intermittent daily process', Hydrology paper 89, CSU,Fort collins,Colorado.

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

40. Klemes,V.(1978),' Physically b ased stochastic hydrologic analysis', Advances in Hydroscience,Vol. 11,pp.285-358,Academic Press, New York,New York.

41. 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.

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

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

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

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

46. Kottegoda,N.T.(1979),' Stochastic water resources technology', John Wiley and Sons, New York.

47. Krzysztofowicz,R.(1985),' Bayesian models of forecasted time series', Water Resources Bulletin Vol.21/5,pp.805-814.

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

49. Kumar,A.(1982),' Hydrologic Time Series Modelling-An overview', State of Art report,SA-1, NIH,Roorkee.

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

51. 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.

52. 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.

53. Lewis,P.A.W.(1985),' Some simple models for continuous variate time series,' Water Resources Bulletin,Vol. 21/4,p.635-644.

54. 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.

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

56. Maissis,A.H.(1977),' Optimal filtering technique for hydrological forecasting', Water Resources Research,Vol. 13,pp.87-98.

57. Mandelbrot,B.B.(1971),' A fast fractional Gaussian noise generator', Water Resources Research,Vol.7/3 pp.543-553.

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

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

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

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

62. Matalas,N.C.(1977),' Generation of multivariate synthetic flows', Mathematical Models for Surface Water Hydrology by Ciriani,Maione and Wallis , John Wiley and Sons,pp.27-38.

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

64. Mc Kenzie,E.(1985),' Some simple models for discrete variate time series', Water Resources Bulletin, Vol.21/4,pp.645-650.

65. Mejia,J.M. Ahlert,R.C. and Yu.S.L.(1975),' Stochastic variation of water quality of the Passaic river', Water Resources Research,Vol.11/2.

65a. Mejia,J.M.(1971),'On the generation of multivariate sequences exhibiting the Hurst phenomenon & some flood frequency analysis',Ph.D. Dissertation,CSU,Fort Collins,Colorado.

66. Mejia,J.M., Rodriguez-Iturbe,I., and Dawdy D.R. (1972),' Streamflow simulation. 2. the broken line process as a potential model for hydrologic simulation; Water Resources Research,Vol.8/4,pp.931-941.

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

68. Mekrecher,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.

69. Moreau,D.H. and Pyatt,E.E.(1970),' Weekly and monthly flows in synthetic hydrology', Water Resources Research,Vol.6,pp.53-61.

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

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

72. Pagano,M.(1978),'On periodic and multiple autoregressions', The annals of statistics,Vol.6/6,pp.1310-1317.

73. 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.

74. Pfeifer,P.E. and Deutsch,S.J.(1980),' Indenfification and interpretation of first order Space Time ARMA models', Tecnometrics,Vol.22,pp.397-408.

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

76. Rao,A.R., Rao,S.G. and Kashyap,R.L.(1985),' Stochastic analysis of time aggregated hydrologic data', Water Resources Bulletin,Vol.21/5, pp.757-772.

77. Rodriguez,I. and Yevjevich,V.(1967)', Sunspots and Hydrologic Time Series', Proceedings of the International Hydrollgy Symp.,Fort Collins,Vol.1,pp.397-405.

78. Roesner,L.A. and Yevjevich, V.(1966),' Mathematical Models for time series of monthly precipitatio n and monthly runoff', Hydrology paper No.15, Colorado State Univ. Fort Collins.

79. Salas,J.D. and Smith,R.A.(1980),' Uncertainties in hydrologic time series analysis', ASCE Spring meeting,portland,Oregon,pp.80-158.

80. Salas,J.D. and Smith,R.A.(1981),' Physical basis of stochastic models of annual flows', Water Resources Research,Vol.17/2, pp.428-430.

81. Salas,J.D.,Delleur,J.W., Yevjevich,V. and Lane,W.L., (1980),' Applied modelling of hydrologic time series Water Resources Publication,Fort Collins,p.484.

82. Salas,J.D. Tabios, G.O. III, and Bartolini, P.(1985),'Approaches to multivariate modelling of water resources time series', Water Resources Bulletin, Vol.21/4,pp.683-708.

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

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

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

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

87. Sudler,C.E.(1927), ' Storage required for the regulation of streamflow', Trans.ASCE,Vol.91,pp.622-660.

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

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

90. Thompstone,R.N. Hipel,K.W., and Mcleod ,A.I.(1985),' Forecasting quarter monthly river flow', Water Resources Bulletin,Vol.21/5,pp.731-742.

91. Tong, H., Thanoon,B. and Gudmundsson,G.(1985),'Threshold time series modelling of two icelandic riverflow systems', Water Resources Bulletin,Vol.21/4,pp.665-661.

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

93. Vecchia,A.V.(1985),' Periodic autoregressive moving average (PARAMA) modelling with applications to water resources', Water Resources Bulletin,Vol.21/5, pp.721-730.

94. Vecchia,A.V., Obeysekra,J.T.,Salas,J.D., and Boes D.C.,(1983),' Aggregation and estimation for low order periodic ARMA models', Water Resources Research,Vol.19/5,pp.1297-1306

95. Weiss,G.(1973a),' Shot noise models for synthetic generation of Multisite daily stramflow data', Symp. UNESCO World Meteorol.Organization,Inter. Association of Hydrol.Sciences,Madrid.

96. Weiss,G.(1973b),' Filtered Poisson processes as models for daily streamflow data', Ph.D. Thesis, Imperial college of Sciences and Technology,London.

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

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

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

100. Yevjevich,V.M.(1963) ,' Fluctuations of wet and dry years,Part 1,Research data assembly and mathematical models', Hydrology paper 1,CSU,Fort Collins ,Colorado.

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

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

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

104. Yevjevich,V. and Harmancioglu,N.B.(1985),' Past and future of analysis of water resources time series', Water Resources Bulletin,Vol.21/4,pp.625-633.

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