RN-25

# RANGE ANALYSIS FOR STORAGE

SATISH CHANDRA
DIRECTOR

STUDY GROUP

P V SEETHAPATHI
S K JAIN

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

1985-86

CONTENTS

# LIST OF FIGURES

## ABSTRACT

The range analysis is a part of the general probability theory of storage. In this report the works dealing with the range analysis for storage related problems have been reviewed. In the beginning, the empirical method and the experimental method are described. This is followed by analytical results for expression of range for a number of processes which are frequently encountered in hydrology.

Extensive experiments on natural time series were conducted by H.E.Hurst. One puzzling result which he obtained was that these series do not behave, in some respects, the way a random series should. This fact was termed 'Hurst Phenomenon'. Several arguments have been advanced by different investigators to explain the Hurst phenomenon; these are briefly discussed. The recognition of Hurst phenomenon has been a turning point in hydrological modelling of time series and this has been briefly discussed. A detailed review of various experiments conducted to determine the effect of inflow generating mechanism on reservoir capacity has been made. At this stage it is not possible to conclusively say about the utility of long or short memory flow models in reservoir analysis because of unavailability of results of exhaustive studies.

# 1.0 INTRODUCTION

Storage reservoirs are one of the most important component of a water resources system. Naturally, the analysis of problems related with storages has drawn considerable attention of hydrologists. The various aspects of this problem which has been studied in considerable detail include the storage required for specified yield, effect of the statistical properties of inflows, such as serial correlation, persistance and range, on the storage requirements and optimal management of single or multiple reservoirs. The solution techniques used for these problems include graphical methods, pure mathematical analysis, simulation and mathematical programming techniques.

## 1.1 Classical Methods of Analysis of Storage

The work of Moran (1959) in which he applied the probability theory to problems of finite storage is considered to be the pioneering effort in mathematical analysis of problems of storage. The study assumed, that the inflows to the reservoir are independent and identically distributed random variables. The water enters the reservoir during wet seasons and is stored and released during the dry seasons. The reservoir has a finite storage capacity and in case the initial storage and inflow in a period exceeds this capacity, extra water over this capacity is spilled. The release from the reservoir is controlled by the standard linear operating policy. Under these assumptions, Moran was able to derive the stationary transition probability matrix of storage.

Langbein(1958) presented the solution of determining the frequency distributions of storage, reservoir outflows and frequency of the reservoir spilling or remaining empty. These solutions were obtained using the queuing theory. Two solutions were given - first was applicable to linear service functions and normal inflows and the second was applicable to nonlinear service functions and non-normal inflows which was based upon probability routing.

The model of Moran was extended by Lloyd (1963) by considering the serial correlations. It was assumed that the inflows follow a homogeneous Markov chain. The limiting distribution of storage was derived using the bivariate Markov process, representing the joint distribution of storage and inflows.

In the fifties, Hurst conducted a series of experiments on a number of time series and gave interesting but puzzling results. Range was the most important parameter studied in these studies. The unusual behaviour of the parameter range which was pointed out by Hurst was later on called 'Hurst phenomenon'. Ever since its recognition, this phenomenon is considered to be turning point in storage analysis and time series modelling.

## 1.2 Scope of the Present Report

In the present report, the published research work dealing with the parameter range is reviewed. The reviewed work include estimation of range, explanations of the Hurst phenomenon, subsequent related development in time series modelling and sensitivity of reservoir storage to the inflow generating mechanism.

## 2.0 DEFINITION OF RANGE AND RELATED TERMS

Let $x_i$, i=1,2,.......N represent a time series of flows at a particular site on a stream. This time series can be a daily, weekly or monthly series. Now assuming that these flows are being fed into a very big reservoir and an amount of water equal to the mean of the series($\bar{x}$) is being taken out.

Let $$S_1 = \Delta x_1 = x_1 - \bar{x} \qquad \ldots(1)$$

where
$$\bar{x} = \Sigma x_i/N$$

represents increase or decrease (depending upon the relative magnitude of $x_1$ & $\bar{x}$)in the reservoir content at the end of first time period.

Similarly,

$$S_2 = \Delta x_1 + \Delta x_2 \qquad \ldots(2)$$

represents this change at the end of the second time period

and $$S_i = \Delta x_1 + \Delta x_2 + \ldots + \Delta x_i = \sum_{j=1}^{i} \Delta x_j \qquad \ldots(3)$$

represents such change at the end of the $i^{th}$ time period. The maximum of n values of $S_i$, denoted by $S_n^+$ is called maximum surplus, or surplus or maximum partial sum of deviates. The minimum value of n values of $S_i$ represented by $S_n^-$ is called minimum deficit or deficit or minimum partial sum of deviations. The sum of magnitude of surplus and deficit, i.e,

$$R_n = S_n^+ + |S_n^-| \qquad \ldots(4)$$

is called the range. The terms surplus, deficit and range are graphically represented in figure 1.

The parameter range is always greater than zero. The range represents the storage capacity required in a reservoir to maintain an outflow equal to $\bar{x}$ if the $x_i$ were inflows to this reservoir. However, this is an ideal requirement which cannot be attained in practice because it involves no spilling or other losses.

The statistic range depends upon the properties of the series as well as upon its size. As the size of the series increases, the range will either increase or will remain the same.

## 2.1 Adjusted Surplus, Deficit and Range

In the above discussion, it was mentioned that the outflow from the reservoir was equal to the mean of all inflows. A different value of the parameters under discussion is obtained if the outflow is not $\bar{x}$ but equal to $\bar{x}_n$, which is the mean of the subseries of size n. In such cases, the adjective adjusted is used with the parameters and they are called adjusted surplus, adjusted deficit and adjusted range. These are graphically depicted in figure 2.

## 2.2 Estimation of Surplus, Deficit and Range

Currently, three approaches are used for determining the parameters surplus, deficit and range. These are:

a) Empirical approach,
b) Experimental approach,
c) Analytical approach,

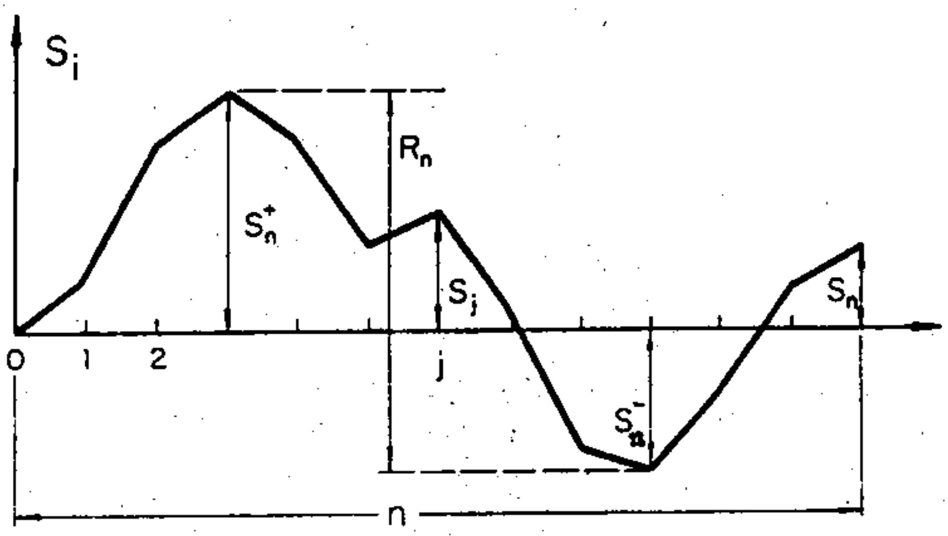


FIGURE 1 - DEFINITION OF SURPLUS,DEFICIT AND RANGE FOR A SAMPLE OR A SUBSERIES OF SIZE n (AFTER YEVJEVICH,1972)

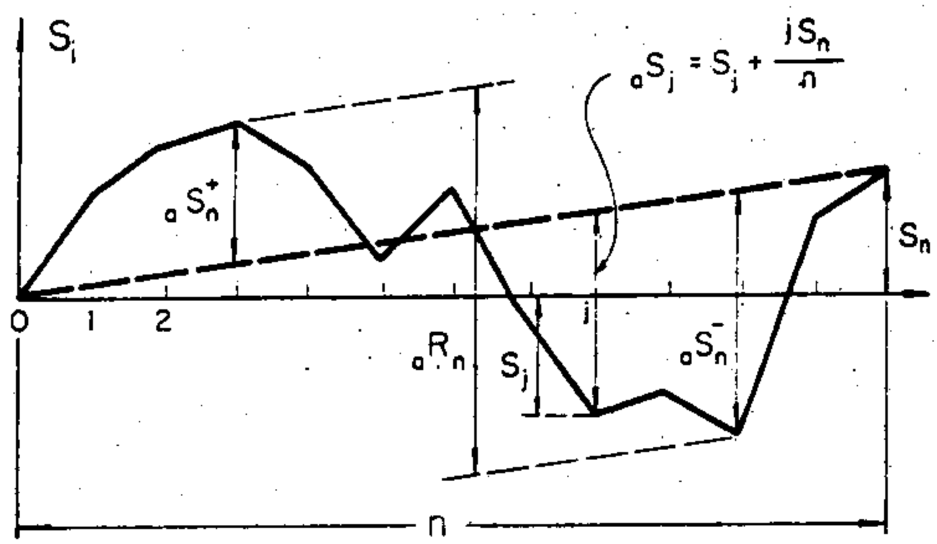


FIGURE 2 - DEFINITION OF ADJUSTED SURPLUS, ADJUSTED DEFICIT AND ADJUSTED RANGE FOR A SAMPLE OF SUBSERIES OF SIZE n (after Yevjevich, 1972)

Here, first two methods are being discussed in detail. The analytical method is only of theoretical interest and can be referred to in standard texts such as Yevjevich(1965).

2.2.1 Empirical Approach

This approach, also known as mass curve method was first developed by Rippl in 1883. In general, the parameters, surplus, deficit and range can be applied to any physical phenomena which can be accumulated in space such as fluid flow,dissolved pollutants in water,dissolved oxygen content in water, soil moisture, sediment discharge and heat etc. Hurst(1964) determined these statistics for a large number of time series such as discharges of a number of rivers including Nile,Thames, Danube,Godavari,Mississippi,Niger,Rhine,Colorado and Missouri, for river levels including Roda Cauge on Nile, and level of Rhine, for rainfall observations at Greenwich,Copenhagen, Frankfurt,Rome,Bangalore,Calcutta,Cherapunji,New York etc. The observation studied by Hurst also include temperatures, atmospheric pressures, annual growth of tree rings, thickness of mud layers in lakes, sunspot numbers and also the phenomena involving human factor such as length of reign of kings of England, ancient Egypt,Pope, prices of wheat, cost of living and number of pennies minted. The results obtained by him are of very interesting nature and shall be discussed later.

For the purpose of storage analysis, let a discrete series of length N of a variable x be available. This discrete series is subdivided into non overlapping samples of size n. Thus the number of new subseries will be

$$m = \frac{N}{n} + r \qquad \ldots(5)$$

where r is the remainder. Defining modular coefficient $K_i$ as

$$K_i = x_i/\overline{x} \qquad \ldots(6)$$

the sum of deviations of $K_i$ from their average value is determined. The deviations are given by

$$\Delta K_i = K_i - \overline{K} \qquad \ldots(7)$$

and the sum of deviation by

$$S_i = \sum_{j=1}^{i} \Delta K_j \qquad \ldots(8)$$

Here $\overline{K}$ is the mean value of $K_i$.

Now, for each series of size n, the value of surplus $S_n^+$, deficit $S_n^-$ and range $R_n$ is determined. Thus there will be m values of each of these parameters which can be used to determine their probability density function and other parameters of distribution.

As n becomes larger, m becomes smaller and the reliability of the method decreases, as pointed out by Yevjevich(1965). Further , the reliability increases with increase in N. In many empirical methods as in Rippls' method, m = 1 and hence the reliability of the results is very less.

2.2.2 Experimental Approach

The methods in which problems are solved by conducting experiments on the generated data are known as experimental methods. These methods are also popularly known as Monte

Carlo methods or simulation methods and are particularly useful in cases where it is difficult to solve the problems by mathematical analysis or the closed form solution is not possible.

One big advantage in use of experimental methods is the that there is no theoretical limit on the size of the sample to be generated. This size is limited only by two practical considerations, the computing facilities or finances available and the degree of accuracy desired.

Nowadays, very efficient routines are available for generation of pseudo random numbers. The numbers can be further transformed to yield the series with the desired properties. Once the series of desired length is available, a further analysis can be done in the manner similar to the empirical method.

It has been pointed out by Yevjevich(1972) that the generation of large sample of hydrologic data from a small sample does not give additional information over what is already available in the small sample. Data generation serves two main purposes : a) it enables full extraction of information already contained in the small sample concerning the statistics surplus, deficit and range, and b) the problems which do not have a closed form solution or those which cannot be solved by empirical or analytical approaches can be solved by this approach. In figure 3, the distribution functions of range are plotted alongwith the results obtained by empirical method. It is readily seen that the distributions obtained by experimental approach are more smooth that those obtained

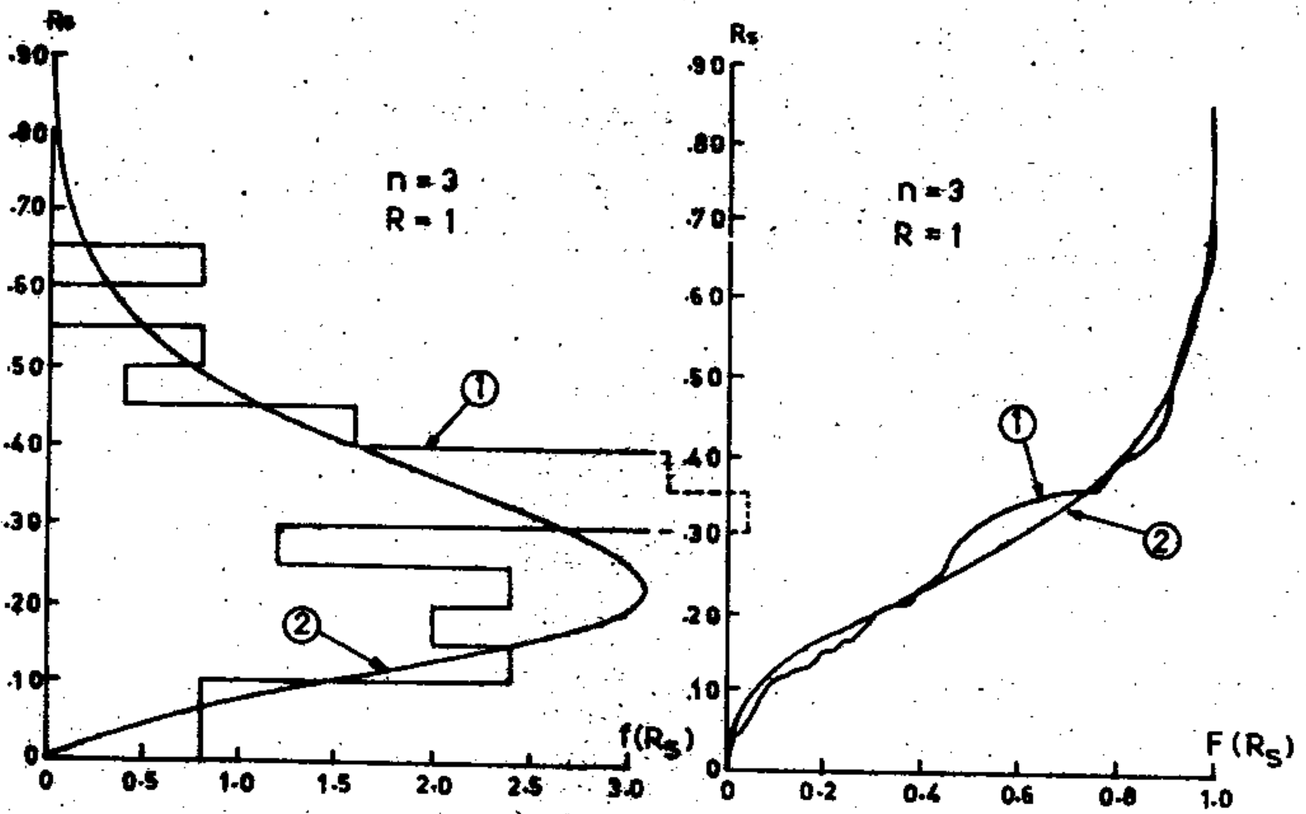


FIGURE 3(a)-FREQUENCY DENSITY OF RANGE $R_3$ OF ANNUAL RIVER FLOWS

FIGURE 3(b)- FREQUENCY DISTRIBUTION OF RANGE $R_3$ OF ANNUAL RIVER FLOWS

1) determined by empirical method,
2) obtained by the data generation method

(After Yevjevich,1965)

by the empirical method. This is considered to be an asset of this method.

A general estimation procedure for Hurst exponent was provided by Mandlebrot and Wallis (1969) through the use of pox diagrams. These diagrams reflect a scatter of points corresponding to the rescaled ranges of full length of a given sample as well as of a set of subsamples extracted from the full sample on the basis of prespecified set of samples sizes. Later on,Wallis and Matalas (1972) employed the least squares procedure for estimating the slope of pox diagrams which leads to an estimation of H.

## 2.3 Range of some processes of interest

In this section, the mathematical expressions for the range statistic for some typical processes, which are commonly encountered in hydrology are given. The derivation of these expressions is not given as they are widely available in texts e.g. Yevjevich(1972),Salas(1972). The following discussion is mainly based on Yevjevich (1972). These results are very useful in understanding the small sample behaviour of the range statistics.

A process which is frequently considered in hydrology is the process described by independent normal variables. The standard normal variable is obtained after substracting mean from the variable and dividing by the standard deviation. Once surplus, deficit and range are available for standard independent normal variables, the same for independent normal variables can be obtained after multiplying by the standard deviation. The expected asymptotic range of an independent normal variable is given by

$$E(R_n) = 2\left[\frac{2n}{\pi}\right]^{1/2} \qquad \ldots(9)$$

The asymptotic adjusted range is given by

$$E\ ({}_aR_n) = \sqrt{\frac{\pi n}{2}} \qquad \ldots(10)$$

The expression for the mean range of any linearly dependent normal variable is

$$E(R_n) = \sqrt{\frac{2}{\pi}}\,\sigma \sum_{i=1}^{n} i^{-1/2} \left[1 + \frac{2}{i} \sum_{k=1}^{i-1} (i-k)\ \rho_k\right]^{1/2} \qquad \ldots(11)$$

where $\rho_k$ is the autocorrelation coefficient of the $k^{th}$ order.

In case of a $m^{th}$ order linear autoregressive process, the autocorrelation coefficient can be expressed as

$$k = \sum_{r=1}^{m} \alpha_r\ \rho_{k-r} \qquad \ldots(12)$$

and the mean range as

$$E(R_n) = \sqrt{\frac{2}{\pi}}\,\sigma \sum_{i=1}^{n} i^{-1/2} \left[1 + \frac{2}{i} \sum_{k=1}^{i-1} (i-k) \sum_{r=1}^{m} \alpha_r\ \rho_{k-r}\right]^{1/2} \qquad \ldots(13)$$

Another type of models, which are also quite common in hydrology are the moving average models. For these models, Yevjevich(1972) gave the expression of range as:

$$E(R_n) = \frac{2}{(m+1)} \sum_{i=1}^{n} i^{-1/2} \left[1 + \frac{2}{i}\ \frac{1}{m+1} \sum_{k=1}^{i-1} (i-k)(m-k+1)\right]^{1/2} \qquad \ldots(14)$$

where m is the order of the process.

Markov models are a class of models which are extensively used in hydrology. For the processes which exhibit Markov behaviour, the expected value of the range is by the empirical method. This is considered to be an asset of this method.

A general estimation procedure for Hurst exponent was provided by Mandlebrot and Wallis (1969) through the use of pox diagrams. These diagrams reflect a scatter of points corresponding to the rescaled ranges of full length of a given sample as well as of a set of subsamples extracted from the full sample on the basis of prespecified set of samples sizes. Later on,Wallis and Matalas (1972) employed the least squares procedure for estimating the slope of pox diagrams which leads to an estimation of H.

2.3 Range of some processes of interest

In this section, the mathematical expressions for the range statistic for some typical processes, which are commonly encountered in hydrology are given. The derivation of these expressions is not given as they are widely available in texts e.g. Yevjevich(1972),Salas(1972). The following discussion is mainly based on Yevjevich (1972). These results are very useful in understanding the small sample behaviour of the range statistics.

A process which is frequently considered in hydrology is the process described by independent normal variables. The standard normal variable is obtained after substracting mean from the variable and dividing by the standard deviation. Once surplus, deficit and range are available for standard independent normal variables, the same for independent normal variables can be obtained after multiplying by the standard deviation. The expected asymptotic range of an independent normal variable is given by

$$E(R_n) = 2\left[\frac{2n}{\pi}\right]^{1/2} \quad \ldots(9)$$

The asymptotic adjusted range is given by

$$E\ ({}_aR_n) = \sqrt{\frac{\pi n}{2}} \quad \ldots(10)$$

The expression for the mean range of any linearly dependent normal variable is

$$E(R_n) = \sqrt{\frac{2}{\pi}}\,\sigma \sum_{i=1}^{n} i^{-1/2} \left[1 + \frac{2}{i} \sum_{k=1}^{i-1} (i-k)\ \rho_k\right]^{1/2} \quad \ldots(11)$$

where $\rho_k$ is the autocorrelation coefficient of the $k^{th}$ order.

In case of a $m^{th}$ order linear autoregressive process, the autocorrelation coefficient can be expressed as

$$k = \sum_{r=1}^{m} \alpha_r\ \rho_{k-r} \quad \ldots(12)$$

and the mean range as

$$E(R_n) = \sqrt{\frac{2}{\pi}}\,\sigma \sum_{i=1}^{n} i^{-1/2} \left[1+\frac{2}{i} \sum_{k=1}^{i-1} (i-k) \sum_{r=1}^{m} \alpha_r\ \rho_{k-r}\right]^{1/2} \quad \ldots(13)$$

Another type of models, which are also quite common in hydrology are the moving average models. For these models, Yevjevich(1972) gave the expression of range as:

$$E(R_n) = \frac{2}{(m+1)} \sum_{i=1}^{n} i^{-1/2} \left[1+\frac{2}{i}\ \frac{1}{m+1} \sum_{k=1}^{i-1} (i-k)(m-k+1)\right]^{1/2} \quad \ldots(14)$$

where m is the order of the process.

Markov models are a class of models which are extensively used in hydrology. For the processes which exhibit Markov behaviour, the expected value of the range is

$$E(R_n) = \sqrt{\frac{2}{\pi}}\,(1-\rho^2)^{1/2} \sum_{i=1}^{n} i^{-1/2} \left[ \frac{1+\rho}{1-\rho} - \frac{2\rho(1-\rho^i)}{i(1-\rho)^2} \right]^{1/2} \quad \ldots\ldots (15)$$

## 3.0 Hurst Phenomenon

Extensive investigation on the properties of range were conducted by Hurst and these have been given in detail by Hurst et.al.(1965). It was concluded that the rescaled range R/ σ , where σ is the standard deviation, increases with the length of the series. It can be proved by mathematical analysis that if a series follows a normal distribution and its members are independent of each other then, for this series for large values of N,

$$\frac{R}{\sigma} = \frac{1}{2}\sqrt{N\pi}$$

$$= 1.25\sqrt{N} \qquad \text{...(16)}$$

However, Hurst found that for the natural phenomenona the relationship between R/σ and the length of the series of given by

$$\frac{R}{\sigma} = \left(\frac{N}{2}\right)^{H} \qquad \text{...(17)}$$

where H is a variable. This equation was derived based on analysis of 75 phenomena and 690 portions of these. It was found that the variable H is normally distributed with mean 0.73 and standard deviation 0.09.

Out of the above two equations foı rescaled range, the first (equation-16) expresses that it increases with 0.5 power of H and the second (equation-17) shows that it increases with 0.73 power of H This discrepancy in the value of exponent H is termed as 'Hurst phenomenon'.

Since the time of this unusual of natural variables was observed by Hurst, tremendous amount of work has been done by different investigators to study the causes of this behaviour and a number of models have been developed to reproduce this phenomenon in the time series models. Before going for a discussion of causes of Hurst phenomenon, the dependance structure of hydrologic time series is being discussed.

A number of studies have been conducted on independent time series but they are of no interest here. The attention is focussed here on the dependance in time series.

## 3.1 Dependence in Hydrologic Time Series

A dependant time series is a series in which any particular element is influenced by its predecessors. In other words, the past history of the series shapes the present. The dependance of a series can be analysed either by correlogram analysis or the range analysis.

Let us assume that the hydrologic time series being studied is stationary. A stationary series is generated by a stationary process whose probability laws do not change with time. Generally the geophysical, biological and other natural processes are assumed to be nonstationary but within a relatively short time spans, they can be assumed stationary.

The hydrologic time series display two types of dependance: long term and short term. If the autocorrelation coefficient of a hydrologic time series is computed, it is observed that it dies out as the lag increases. This implies that a particular value of the variable is influenced only by the recent past values of the series, the distant values do not affect the present. This type of dependence is termed as short term dependance and the process is said to be a short memory process. In these processes the incidents tend to fade from the system memory as the time passes. Although this also looks intutively correct, this observation fails to explain Hurst phenomenon.

The Hurst phenomenon can be satisfactorily explained by long term dependence , which implies that the process has infinite memory. The long term dependence is associated with failure of the correlogram to die at higher lags. However, it is difficult to explain this physically, for example, it is hard to figure out as to how the discharge of a particular day is affected by the discharge of say 100 past days and by what mechanics, the impact of a hydrological event is carried over for years together. These counter arguments have given rise to controversy about the appropriate explanation. The short term dependence, though physically believable, cannot explain the Hurst phenomenon. On the other hand, the long term dependence can explain this feature but it can not be physically explained.

Two types of models have been developed in hydrology corresponding to these two types of dependences: the long memory models and the short memory models. The causes of Hurst phenomenon will be discussed in the next section and this will be followed by discussion on these model types.

### 3.2 The Causes of Hurst Phenomenon

Wallis and Matalas(1970) have given four causes which explain why the time series exhibit Hurst phenomenon. These causes are discussed below:

It has been suggested by various investigators that the non normality of the probability distributions of the time series may be the cause of the Hurst phenomenon. However, the simulation studies carried on using the samples of moderate lengths show that $R_n^*$, the rescaled adjusted range, is quite independent of the distribution of random variable. Hence this cause may really not be able to explain the Hurst phenomenon.

It was pointed out by Hurst that the high values of the exponent H can be attributed to the non-stationarity of the observed series. A stationary process is a process whose probability laws do not change with time. According to Klemes (1974), the geophysical, biological, economic and other natural processes are nonstationary but within a relatively short time span, they can be approximated by stationary models. Thus a longer series will have more chances of being nonstationary. By the assumption of stationarity of process, it is implied that the driving parameters of the process have been fixed in the beginning and they remain unchanged later on. On the other hand, for a nonstationary process, the parameters do change with the time and respond to the dynamics of the system and environment.

It has been argued by Klemes (1974) that stationarity in the mean is the most important prerequisite for sound interpretation of a correlogram because the mean represents the absolute reference from which the deviations are measured. Thus stationarity is also a pre-requisite for the sound interpretation of Hurst phenomenon since the rescaled range is also a function of deviations from the mean. Klemes(1974) conducted a number of simulations using white noise and the mean level was changed in different manners. It was shown that H increases with this type of nonstationarity. The nonstationarity assumption may not be very helpful in practice since it is very difficult to fit nonstationary model to a given hydrologic series.

Another explanation which was put forward to explain Hurst phenomenon was that the length of available records is not long enough for H to attain a value of 0.5. It has been argued that if a sufficiently long series of observation is available, H would tend to attain a value 0.5. At this stage, this argument can neither be accepted nor rejected

A plausible explanation of the value of H higher than 0.5 is the persistence, the higher values being the effect of dependence in observed natural series. The dependence can be taken into account in Markovian models but these models cannot reproduce $H > 0.5$.

Hence, if the dependence is to be considered as the likely cause of Hurst phenomenon than a different type of models, called long memory models are required. The short term persistence is caused by the storage effect.

Kumar (1982) pointed that in the geophysical processes, the memory manifests itself mostly through the conservation of mass and momentum and it has the Markovian property that the past influences the future, only through its influence on the present. Thus, once the present state has been arrived at, it is no longer significant from the point of view of future development as to how it was arrived.

Based upon the above discussion, from the point of view of interest, all the available time series models can be classified into two groups - short memory models and long memory models.

## 3.3 Short Memory Models

The main use of short memory models in hydrology has been to generate synthetic data sequences. For many hydrologic studies, particularly those concerned with the design and management of water resources system, sufficiently long data series are required for the purpose of determining the operating rules and testing the system under various possible conditions to evaluate its performances. It may be mentioned that a longer series does not contain any further information than the original series which was used to generate it but it helps in greater extraction of the information already contained in the original series. The essential requirements of a generated series is that it maintains the statistical properties of the original series.

One type of the short memory models which have been extensively used in hydrology are autoregressive models. In these models the current value of the process is expressed as a finite linear aggregate of previous values of the process and noise term. Let $Z_t$ by a typical process under consideration and $\mu$ be the mean of the process. The model works with the deviations of the process from the mean

$$Z_{dt} = Z_t - \mu \qquad ...(18)$$

Mathematically, the process is represented as

$$Z_{dt} = \phi_1 Z_{dt-1} + \phi_2 Z_{dt-2} + .... \phi_p Z_{dt-p} + a_t \qquad ...(19)$$

This process is an autoregressive process of order p, also represented as AR( p) process. It has (p+2) parameters: $\mu$ , $\phi_1$, $\phi_2$, ....$\phi_p$ and $\sigma_a^2$ which is the variance of the random variate $a_t$.

Another type of models frequently used are the moving average models. Here the deviation of the process from mean is expressed as a finite weighted sum of random elements (white noise) plus a random element:

$$Z_{dt} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} .... \theta_q a_{t-q} \qquad ...(20)$$

where $\theta_i$ are weights. This equation represents a moving average process of the order or a MA (q) process. It has (q+2) parameters:

$\mu, \theta_1, \theta_2$ ..... $\theta_q$, and $\sigma_a^2$ which is the variance of $a_t$.

Before applying autoregressive and moving average models to any problem, it is necessary to find the order of the model, i.e., values of parameters p and q respectively. During the development of a stochastic model, one should go for an adequate but parsimonious model, since a model with unnecessary parameters can lead to poor results. Further, many times, either autoregressive or moving average models may not be able to faithfully represent the process. In the composite model as described by Box and Jenkins(1976), the deviation of a process from its mean is expressed as a finite weighted sum of p previous deviations plus a finite weighted sum of q random variates and a random element. Hence an ARMA(p,q) model having p autoregressive terms and q moving average terms would look like:

$$z_{dt} = \phi_1 z_{dt-1} + \phi_2 z_{dt-2} + \ldots\ldots\ldots \phi_p z_{dt-p}$$

$$+ a_t - \theta_1 a_{t-1} \ldots\ldots\ldots\ldots \theta_q a_{t-q} \qquad \ldots(21)$$

Thus an ARMA (1,1) model would be

$$z_{dt} = \phi_1 \; z_{dt-1} \; a_t - \theta_1 \; a_{t-1}$$

This model has been extensively used in hydrology, both for data generation and for forecasting.

The presence of seasonal cycles and trends in hydrologic data causes certain problems in analysis. These can be overcome by differencing the series, i.e., by substracting a particular value from its previous value or its value j units apart. The models in which this exercise is done before fitting an ARMA model are called Autoregressive Integrated Moving Average models or ARIMA(p,d,q) models. Here d denotes the order of the differencing.

### 3.4 Long Memory Models

The purpose of long memory models is to model long term dependence which is caused by the presence of a low frequency component. In practice, this is typified by long periods of either very high flows or very low flows. Few time series models which have been developed to model this effect are described here in brief.

3.4.1 Fractional Gaussian Noise Model

The fractional Gaussian noise (f Gn) model was developed by Mandelbrot and Van Ness (1968). An important component of fGn models is the Brownian motion process, which is a stochastic process. If B(t) is a Brownian motion process then its increments B (t+a)-B (t) are Gaussian with mean equal to zero and variance equal to u and these are independent for non overlapping time intervals. A fractional Brownian motion can be defined as the moving average of the incremental continuous time process dB(t) = B(t+dt)-B(t) in which past increments of B(t),dB(s), are weighted by $(t-s)^{H-0.5}$. Here H is the Hurst's exponent which varies between 0 .5 and 1.

Since modelling of these processes in continuous time is extremely complex, they are always dealt with in discrete time. A discrete fractional Gaussian noise (dfGn) is a Gaussian random process with a $k^{th}$ order autocorrelation coefficient given by

$$\rho(k) = \frac{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}}{2} \quad \ldots(23)$$

In practice approximations of dfGn are used because to construct a sample function of dfGn, and infinite number of components have to be summed up. However, the number of terms must be large enough to preserve the required value of Hurst exponent.

3.4.2 Fast Fractional Gaussian Noise Model

The fast fractional Gaussian noise (dfGn) models have been introduced to overcome the computational difficulties of fGn. The models are discrete approximations to theoretical fractional Gaussian noise. A ffGn model is made up of three components:

i) an independent autoregressive process. This gives the high frequency effects:

ii) low frequency term which reproduces the low frequency properties of the covariance function, and

iii) a random element.

De Coursey et al.(1982) have described these models in sufficient details and have concluded that the ffGn is the best of the discrete fGn models.

In an effort to compare the long memory and short memory models, Lettenmaier and Burges(1977) compared reservoir storage requirements using ffGn and ARMA(1,1) models. The value of Hurst coefficient was assumed to be 0.7. In case of life of reservoir as 40 years, both models gave more or less same results. But for reservoir life of 100 years, the results showed significant difference. However, when they combined the low frequency ARMA(1,1) model with high frequency lag-1 Morkov process, the results were again same.

## 3.5 Markov Models

A Markov process has the property that the current value of the process depends only on the event of immediate past. Thus the effect of past history on the future is manifested only through the present value. Mathematically, if x(t) follows a Markov process then

$$F\left[x(t+k)/x(t),x(t-1),\ldots\ldots\right] = F\left[x(t+k)/x(t)\right] \quad k > 0 \quad \ldots(24)$$

Since the current value completely defines the state of a Markov process, this value is also referred as the 'state' of the process.

A Markov process which can only take the discrete values is called Markov Chain. The hydrologic processes such as stream flows, reservoir storages have to be almost always discretized for ease of computations. The properties of Markov chains are described here with the help of an example.

Let $Q_y$ be the annual streamflow of a particular river in year y. It is assumed that the distribution of these random variables is stationary. If this random variable takes on values $q_i$ with probability $p_i$ then

$$\sum_{i=1}^{n} p_i = 1 \qquad \ldots(25)$$

where n is the number of discretized states. The dependence of $Q_{y+1}$ on $Q_y$( assuming Markov Property) can be specified by transition probabilities which specify the transition of the variable from a given state in a time period to another state in the next time period. Mathematically,

$$P_{ij} = \Pr\left[Q_{y+1} = j/Q_y = i\right] \qquad \ldots(26)$$

where $P_{ij}$ is the transition probability from state i to state j. This is the conditional probability that the next state is j given the present state i. Naturally they should satisfy

$$\sum_{j=1}^{n} P_{ij} = 1 \quad \text{for all i values} \qquad \ldots(27)$$

Generally these probabilities are presented in the form of a matrix called transition probability matrix. This matrix completely describes a Markov chain. Let $p_i^y$ be the probability that the system is in state i in a year y then the probability of $p_j^{y+1}$ , i.e., system is in state j in year(y+1) is

$$p_j^{y+1} = p_1^y \; p_{1j} + p_2^y \; p_{2j} \ldots p_n^y \, p_{nj}$$

$$= \sum_{i=1}^{n} p_i^y \; p_{ij} \qquad \ldots(28)$$

If p represents the transition probability matrix then

$$p_j^{y+1} = p_j^y \; p \qquad \ldots(29)$$

Similarly, by induction,

$$p^{y+2} = p^{y+1} \, p$$

$$= (p^y p) \; p$$

$$= p^y p^2$$

Hence $$p^{y+k} = p^y \; p^k \qquad \ldots(30)$$

It can be shown that as the time progresses the probabilities reach limiting values. These limiting probabilities are called steady state transition probabilities.

The recent developments in the above model types have been discussed in detail by Hipel et.al.(1977).

In a number of attempts to explain the Hurst phenomenon, it has been assumed that the natural time series can be represented by a Markov process. But the operational utility of Markov processes in water resources studies is dependent upon the availability of estimates of low order moments, including several lag and cross correlation coefficients which describe the inflows. It has been shown by various investigators that unless the sequences are sufficiently long by hydrologic standards, the reliability of estimates of these coefficient will be unacceptable.

As an alternatives to Markov processes, the fractional Brownian processes were introduced by Mandelbrot and Wallis (1969). Regardless of their order, for the Markov processes, the value of Hurst coefficient asymptotically coverages to 1/2 as N increases whereas for fractional Brownian processes, $0 \leq H \leq 0.5$ or $0.5 \leq H \leq 1.0$. Further a particular value of H remains constant for all values of N.

The value of H varies from stream to stream and consequently, while assessing the reliability of an estimate of H one must take into account the sampling errors, spatial variability, and properties and assumptionsof the method employed.

Wallis and Matalas (1970) made an effort to regionalise H by relating H to a set of basin characteristics. However, their results displayed a large variability and there was no perceptible regional pattern. The authors adopted two techniques to estimate H. The first method was that suggested by Hurst in which the coefficient is estimated by

$$h_H = \log (R/S) \,/\, \log (N/2) \qquad \ldots(31)$$

The second method was suggested by Mandelbrot and Wallis (1969) in which the time series of N observations is subdivided into smaller sets of length n and for each of theses subseries, R(n) and S(n) are determined. The value of h,denoted by $h_M$( just for the purpose of distinguishing) is a given by the slope of line log $[R(n) / S(n)]$ versus log n.

Based upon a number of simulation experiments, Wallis and Matalas (1970) concluded that the distribution properties of $h_H$ and $h_M$ are function of the sequence length N and the probability distribution underlying the sequence values. The distributions of $h_H$ and $h_M$ were positively skewed and they observed that the skewness decreases with the increase in N. Further both $h_H$ and $h_m$ were found to be biased estimators of H and the bias for $h_H$ was larger than the bias for $h_M$. However, the variance of $h_H$ is less than the variance of $h_M$. The variance for $h_M$ was found to be a function of $n_o$ which is the lower limit of the values of n over which h was determined. Interestingly, the bias of $h_M$ decreases with increase in $n_0$ but the variance increases. For lag one autocorrelation $\rho > 0$, the biases in $h_H$ and $h_M$ increase as $\rho$ increases while for $\rho < 0$, the biases decrease as the absolute value of $\rho$ increases.

The authors, after analysing records of 25 streams in Potomac basin in USA, concluded the following:

a) It would be very difficult to use transience, i.e, N not being sufficiently large for H to attain its limiting value, as the explanation of Hurst phenomenon.

b) Care must be exercised if H is estimated from annual values only.

c) It might be difficult to find monthly Markov models that accurately mime the stramflow regime observed in Potomac basin.

## 3.6 Senstivity of Reservoir Capacity to the Inflow Generating Mechanism

The minimum reservoir capacity required for meeting the given demands depends upon the generating mechanism of inflows apart from the nature of demands themselves. Wallis and Matalas (1972) conducted simulation experiments to determine the sensitivity of the reservoir capacity to various parameters of an inflow sequence including the Hurst exponent H. The authors used two approaches for generation of inflows the Markov process and the fractional Gaussian noise process. The most important parameter of Markovian process is $\rho_u$ which is is lag u serial correlation coefficient. The fractional Gaussian noise model was proposed by Mandelbrot and Wallis (1968); in the above study, a type 2 model was used. For this process, the governing parameters are the Hurst exponent H, and the memory of the process M. It may be mentioned that for Markovian models, H is always 0.5.

The authors generated a number of sequences using these two models and the required reservoir capacity for various levels of development was determined using Sequent Peak Algorithm. Obviously the required storage is a function of the level of development $\alpha$. It was found out that over certain ranges of values of $\alpha$ , $\rho_u$ and h, the required reservoir capacity is insensitive to the inflow generating process. For $\alpha > 0.80$, the capacity depends upon $\rho_u$ and the Markovian models may be confidently used although the filtered Gaussian noise type 2 process gives better representation of the real world. For $\alpha > 0.80$, the capacity mainly depends upon H.

The theoretical value of long memory models lie in their ability to generate time series which resemble long historic records better than those generated by short memory models. Klemes et al. (1981) conducted simulation experiments to determine differences in reservoir performance reliability when the inflows are generated using long memory models and short memory models. The reliability of the reservoir was characterized in following three different ways:

a) Occurrence - based reliability $R_a$ which is the number of nonfailure years expressed as a percentage of total number of years in a given period.

b) Time-based reliability $R_t$ which is the total duration time of all nonfailure intervals expressed as a percentage of total length of given period.

c) Quantity- based reliability $R_v$ which is the actual amount of water supplied expressed as a percentage of the total demand during the given period.

Usually at least some years contain shorter or longer periods of nonfailure and during most failure periods, the outflow is not reduced to zero. This leads to the condition that $R_a \leq R_t \leq R_v$ ...(32)

Based on simulation experinents, the following observations were made by Klemes et al.(1981).

1) The short-memory model leads to over estimation of reservoir performance reliability as compared to the long memory model.

2) The overestimation is, in general, highest for annual reliability, lower for time based reliability and lowest for quantity based reliability.

3) The over estimation of all the three reliability

characteristics is very small for reservoirs with storage coefficients upto about one for any draft ratio; for reservoirs with storage coefficient greater than one, it is very small for draft ratios upto about D = 0.6 to 0.8 and over D = 1.1, depending on inflow parameters like coefficient of variation and lag-one serial correlation coefficient.

4) The over-estimation is maximal for draft ratios close to one and increases with reservoir coefficients upto about 2 or 3. For higher values of this coefficient no further increase was detected.

5) The over-estimation slightly increases with the variability of inflows and decreases with the increase of the lag 1 correlation coefficient.

Thus it was concluded that the use of short or long memory model makes no difference in reservoir performance reliability if the draft ratio is either very low or very high. The length of memory in the annual inflow series is irrelevant if the regulation itself has no over year memory, for example, if the draft is low, the reservoir fills up every year and if it is high, the reservoir empties every year; in both cases, only seasonal regulation is involved.

It was further pointed out that the decision makers with high-risk aversion will prefer the long-memory models;those with low-risk aversion, the short-memory models. The replacement of a short-memory streamflow model with a long-memory model amounts to the incorporation of a small safety factor into the reservoir performance reliability. However, in most practical cases, this factor will be much smaller than the accuracy with which the performance reliability can be assessed.

## 4.0 CONCLUSIONS

The works related with analysis of hydrologic time series pertaining to range have been reviewed. The different techniques used to estimate range of a series have been discussed. The experiments of H.E. Hurst and the Hurst phenomenon have been discussed. Various investigators have proposed different explanations for Hurst phenomenon. Among many arguments, nonstationarity and persistence appear to have maximum weight behind them.

The Hurst phenomenon has greatly affected the developments in time series modelling. These have also been discussed in brief. Finally, the effect of inflow generating mechanism on the storage is considered. Based on the reviewed work, it could be concluded that the use of short memory or long memory models makes no difference in reservoir performance reliability if the draft is either too high or too low. In most of the cases, the difference in safety factor will be much smaller than the accuracy with which the performance reliability could be assessed. However, it may be pointed out that the studied reports have not covered such aspects as effect of operating policies, shorter time period of analysis etc. Unless detailed studies are conducted considering these aspects also, it will not be possible to reach to a definite conclusion.

REFERENCES


1. Bras,R.L., and I.Rodriguez-Iturbe, Random Functions and Hydrology,Addison-Wesley Publishing Company Reading ,1985.

2. Burges,S.J., and R.K.Linsley," Some factors influencins required reservoir storage", Journal of the Hydraulics Division,ASCE,97(HY7),977-1971.

3. DeCoursey,D.G.,et.al.," Stochastic models in hydrology", Chapter 2 in Hydrologic Models of Small Watersheds, C.T.Haan, et al.(ed). Monograph #5 ,ASAE,1982.

4. Fiering,M.B.,Streamflow Synthesis,Harward University Press,Cambridge,Mass,1967.

5. Hipel,K.W.,A.I. McLeod and E.A.McBean," Stochastic modeling of the effects of reservoir operation', Journal of Hydrology,Vol.32,97-113,1977.

6. Hurst,H.E.,R.P. Black, and Y.M. Simaika,Long Term Storage, Constable,London,1965.

7. Klemes,V.,R.Srikanthan, and T.A.McMahon ," Lons-memory flow models in reservoir analysis: What is their practical value ? , Water Resources Research, 17(3), 737-751,1981.

8. Kumar,A.,Hydrologic Time Series Modelling-An Overview, Report No.SA-1, National Institute of Hydrology, Roorkee,1982.

9. Lettenmaier,D.P., and S.J.Burges,' Operational assessment of hydrologic models of long term persistence", Water Resources Research,13(1),113-124,1977.

10. Mandlebrot,B.B., and J.R.Wallis," Noah,Joseph, and operational hydrology', Water Resources Research, 4(5),909-918,1968.

11. Mandlebrot,B.B., and J.R.Wallis," Computer experiment with fractional Gaussian noises,1. Averages and variances', Water Resources Research,5(1),228-241, 1969.

12. Mandlebrot,B.B., and J.R.Wallis," Computer experiments with fractional Gaussian noises, 2. Rescaled ranges and spectra', Water Resources Research,5(1),242-259,1969.

13, Mandlebrot,B.B.," A fast fractional Gaussian noise generator", Water Resources Research,7(3),543-553,1971.

14. McLeod,A., and K.W.Hipel, and W.C.Lenox," Advances in Box-Jankins modelings 2.Applications ", Water Resources Research,13(3),577-586,1977.

15. McLeod ,A., and K.W.Hipel," Preservation of rescaled adjusted range 1. A ressessment of the Hurst Phenomenon", Water Resources Research,14(3),491-507,1978.

16. Potter,K.W., " Evidence for nonstationarity as a physical explanation of the Hurst phenomenon", Water Resources Research,12(5),1047-1052,1976.

17. Sen,Z., " The small sample estimation of h", Water Resources Research,13(4),971-974,1977.

18. Sen,Z., " The small sample expectation of population and adjusted ranges ", Water Resources Research, 975-980,1977.

19. Sen,Z., " The small sample expectation of rescaled population and rescaled adjusted ranges", Water Resources Research,981-986.

20 Wallis,J.R., and N.C. Matalas," Small sample properties of H and K - Estimators of the Hurst Coefficient h", Water Resources Research,6(6),1583-1594,1970.

21. Wallis,J.R., and N.C.Matalas," Sensitivity of reservoir design to the generatins mechanism of inflows", Water Resources Research,8(3),634-641,1972.

22. Yevjevich,V.M., the Application of surplus,deficit, and range in hydrology, Hydrology Paper No.10,Colorado State University,Fort Collins,1965.

23. Yevjevich,V.M., Probability and Statistics in Hydrology, Water Resources Publication,Fort Collins,1972.

24. Yevjevich,V.M.,Stochastic Processes in Hydrology,Water Resources Publication,Fort Collins,1972.

25. Klemes,V.," The Hurst Phenomenon: A puzzle ?,' Water Resources Research 10(4),675-688,1974.

26. Langbein,W.B.," Queing theory and water storage', Proceedings of ASCE,83(HY5),Paper No.1811,1958.

27. LLoyd,E.H.," A probability theory of reservoirs with serial correrelated inputs", Journal of Hydrology,Vol.1,99-128,1963.

28. Moran,P.A.P., The Theory of Storage," Matheun,London,1959.

29. Salas,J.D.,Range Analysis for Storage Problems of Periodic-Stochastic Processes Hydrology Paper No.57,Colorado State University,Fort Collins,1972.