RR-75-29

Reprint

ANALYTICAL STUDIES OF THE HURST EFFECT:
A SURVEY OF THE PRESENT POSITION

A. A. Anis

E. H. Lloyd

August 1975

Analytical Studies of the Hurst Effect: A Survey of the Present Position*

A. A. Anis** and E. H. Lloyd***

The Hydrological Significance of Hurst's Law

Two of the most striking statistical features of hydrology are 1) the general rarity of data, and 2) the complexity required of even moderately realistic models. (In particular, whilst in many branches of engineering practical use can be made of simple stochastic models involving mutually independent random variables, this is almost never the case in hydrology; the variables concerned usually exhibiting complex cross and auto-correlation structures.)

These factors amongst others have led to the widespread use of numerical simulation methods, in which "synthetic data" are generated in large quantities and subjected to numerical manipulation. The generating process must reproduce data as good as the data from what are regarded as the most important features of the historical record. In the twenties of this century these were taken to be the seasonal averages and the daily, monthly, quarterly or annual fluctuations. Later, this list was supplemented by imposing a simple serial correlation structure on the inflows in accordance with what has become standard practice in statistical time-series analysis. More recently the situation has been further transformed by the discovery of the Hurst effect, and research in simulated data generation is now concentrated on methods of producing number sequences showing Hurst-like behaviour

---

*This report forms an expanded version of a seminar presented by the authors at IIASA in May 1975.

**Ain Shams University, Cairo, Egypt.

***University of Lancaster, UK.

It is therefore a matter of interest in hydrology to attempt to understand what the Hurst effect is, and to construct mathematical models for data-generation that are compatible with Hurst's law.

In the work reported on in this paper, we reconsider the interpretation of Hurst's data and the formulation of Hurst's law, propose a theoretical model whose statistical properties can in principle be derived analytically, summarise such analytical results as are available, and indicate a direction of of possible progress.

1. The Adjusted Range

Denote the consecutive annual flows into a reservoir over a period by

$$x_1, x_2, \ldots, x_n \quad ,$$

the initial contents by a, and the consecutive annual abstractions over the same period by

$$w_1, w_2, \ldots, w_n \quad .$$

The reservoir and its initial contents being supposed sufficiently large, the successive net annual contents at the end of each water year will be

$$a,\ a + x_1 - w_1,\ a + x_1 + x_2 - w_1 - w_2, \ldots, a + x_1 + \cdots + x_n - w_1 - \cdots w_n \quad . \tag{1}$$

Denote the largest of these quantitites by $a + m_n$ (where $m_n \geq 0$) and the smallest by $a + \ell n$ (where $\ell n \geq 0$). Then if we neglect fluctuations of levels within each water year and consider only the levels at the end of each year, namely, the quantities (1), the critical conditions on the reservoir to avoid a) spilling and b) completely emptying, for this set of data, are

a) reservoir capacity = $a + m_n$

b) $a + n = 0$;

that is, the required reservoir capacity is

$$m_n - \ell_n = \max(a, a + x_1 - w_1, a + x_1 + x_2 - w_1 - w_2, \dots, a + x_1 + \dots + x_n - w_1 - \dots - w_n)$$

$$= -\min(a, a + x_1 - w_1, a + x_1 + x_2 - w_1 - w_2, \dots, a + x_1 + \dots + x_n - w_1 - \dots - w_n) \qquad (2)$$

$$= \max(0, x_1 - w_1, x_1 + x_2 - w_1 - w_2, \dots, x_1 + \dots + x_n - w_1 - \dots - w_n)$$

$$= -\min(0, x_1 - w_1, x_1 + x_2 - w_1 - w_2, \dots, x_1 + \dots + x_n - w_1 - \dots - w_n)$$

where the notation "max(···)" denotes the largest of the quantities in parentheses, and "min(···)" the smallest. The qantity $m_n - {}_n$ defined in (2) in the range of the accumulated sums of the numbers

$$0,\ x_1 - w_1,\ x_2 - w_2, \dots,\ x_n - w_n \ .$$

In hydrological contexts of this kind the explicit reference to accumulated sums is usually omitted, and one speaks simply of the range.

The ways in which the statistical properties of this range depend on the duration of the record are, clearly, relevant to the design capacity of the reservoir. This is especially true of the mean valu of $m_n - \ell_n$ as a function of n. In studies of this subject the simplification is often made of taking the abstracted quantitites $w_1, w_2, \dots, w_n$ to have a common value w, so that

$$m_n - \ell_n = \max(0, x_1 - w,\ x_1 + x_2 - 2w, \dots,\ x_1 + \dots + x_n - n_w)$$
$$-\min(0, x_1 - w,\ x_1 + x_2 - 2w, \dots,\ x_1 + \dots + x_n - n_w)$$

this being the range of accumulated sums of the quantities

$$0,\ x_1 - w,\ x_2 - w, \ldots, x_n - w \quad .$$

Finally, it is useful to consider the special case when the final contents of the reservoir, at the end of the n-year period, exactly equal the initial contents. In this case

$$x_1 + x_2 + \cdots + x_n - n_w = 0 \quad ,$$

so that

$$w = (x_1 + \cdots + x_n)/n$$

$$= \bar{x}_n \quad , \quad \text{say.}$$

The range $m_n - \ell_n$ for this case is called the <u>adjusted range</u>, r_n^*, to distinguish it from other cases. Thus the adjusted range is

$$r_n^* = m_n - \ell_n$$

$$= \max(0,\ x_1 - \bar{x}_n,\ x_1 + x_2 - 2\bar{x}_n, \ldots,$$
$$x_1 + \cdots + x_{n-1} - (n-1)\bar{x}_n,$$
$$x_1 + \cdots + x_n - n\bar{x}_n)$$

$$-\min(0,\ x_1 - \bar{x}_n,\ x_1 + x_2 - 2\bar{x}_n, \ldots,$$
$$x_1 + \cdots + x_{n-1} - (n-1)\bar{x}_n,$$
$$x_1 + \cdots + x_n - n\bar{x}_n)$$

$$= \max(0,\ x_1 - \bar{x}_n,\ x_1 + x_2 - 2\bar{x}_n, \ldots,$$
$$x_1 + \cdots + x_{n-1} - (n-1)\bar{x}_n)$$

$$-\min(0,\ x_1 - \bar{x}_n,\ x_1 + x_2 - 2\bar{x}_n, \ldots,$$
$$x_1 + \cdots + x_{n-1} - (n-1)\bar{x}_n)$$

since

$$x_1 + \cdots + x_n - x_n - n\bar{x}_n = 0 \quad . \qquad (3)$$

2. Scaling: The Hurst Range and the Hurst Phenomenon

The magnitude of the adjusted range of an n-year record as defined in (3) is, clearly, related to the inherent variability of the data. Highly variable data will usually possess a large adjusted range, whilst relatively invariable data will possess only a small adjusted range. In order to allow for comparisons between different runs of data from a given river, or between sets of data from different rivers, Hurst introduced the idea of scaling the adjusted range $m - \ell_n$ of a given set of data by dividing by the sample standard deviation d_n of the n inflows $x_1, x_2, \ldots, x_n$. The resulting ration

$$r_n^{**} = r_n^*/d_n \qquad (4) \quad .$$

where

$$d_n^2 = \sum_{j=1}^{n} (x_j - \bar{x}_n)^2/n \qquad (5)$$

is called the rescaled adjusted range, or the Hurst range. The Hurst range is of course a non-dimensional quantity, and its numerical value is not affected by the units used in measuring the flows x_j.

The scaling procedure also has a stabilizing effect: some d_n is positively correlated with r_n^*, the ratio $r_n^{**} = r_n^*/d_n$ and has a smaller sampling variability than the unscaled range. (Possible variants in the interpretation of Hurst's range are discussed in Section 3.)

On the basis of an exceptionally large body of data obtained from a wide variety of rivers (and other sources of geophysically equivalent data) Hurst announced in 1954 that the way in which r_n^{**} increased with n was not proportional to $n^{\frac{1}{2}}$, as elementary theory would lead one to expect, but to $n^{0.72}$, or rather to n^h, where h was near to 0.72 in all cases, with relatively small fluctuations from one set of data to another. Explicitly, his formulation was equivalent to:

$$r_n^{**} = (\tfrac{1}{2}n)^h \quad , \qquad (6)$$

where h had a mean value of about 0.72 and a standard deviation of about 0.09. (Our notation differs from Hurst's for reasons which will be explained in a subsequent paper.)

In this paper we refer to the exponent h as the Hurst exponent, the discrepancy between the empirical value of h(=0.7), and the value to be expected on elementary theory (=0.5) as the Hurst effect or the Hurst phenomenon, and the formulation (6) as a version of Hurst's law. (This nomenclature has been created by hydrologists and probabilists, and not by Hurst himself.)

3. Some Possible Ambiguieties in Hurst's Data and His Treatment of It

3.1 The Data

The following table is a brief extract from some of Hurst's data as summarized in Hurst [12]. The summary consists of single rows of data for certain rivers (e.g. the Mississippi in our excerpt, Table 1.) and several rows, corresponding to different but possibly overlapping intervals, in the case of other rivers (such as the Nile at Aswan in our excerpt). Each run is, however, too concise a summary to enable one to see the details. Some light is thrown on these by graphs (given in the same publication), such as the following Figure 1.

Table 1. Accumulated departures, river discharges.

River	Period	Duration n	Std. Devn. d_n	Adj. range r_n^*	Hurst range r_n^{**}	Hurst exponent h
Mississippi	1874/1936	63	13	190	14.6	0.77
Nile	1870/1975	21	13.4	98	7.3	0.74
(Aswan)	1899/1957	59	12.2	70	5.7	0.50
	1870/1913	44	19.3	292	15.1	0.88
	1914/1957	44	11.0	82	7.5	0.65
	1870/1957	88	17.5	500	34.2	0.88

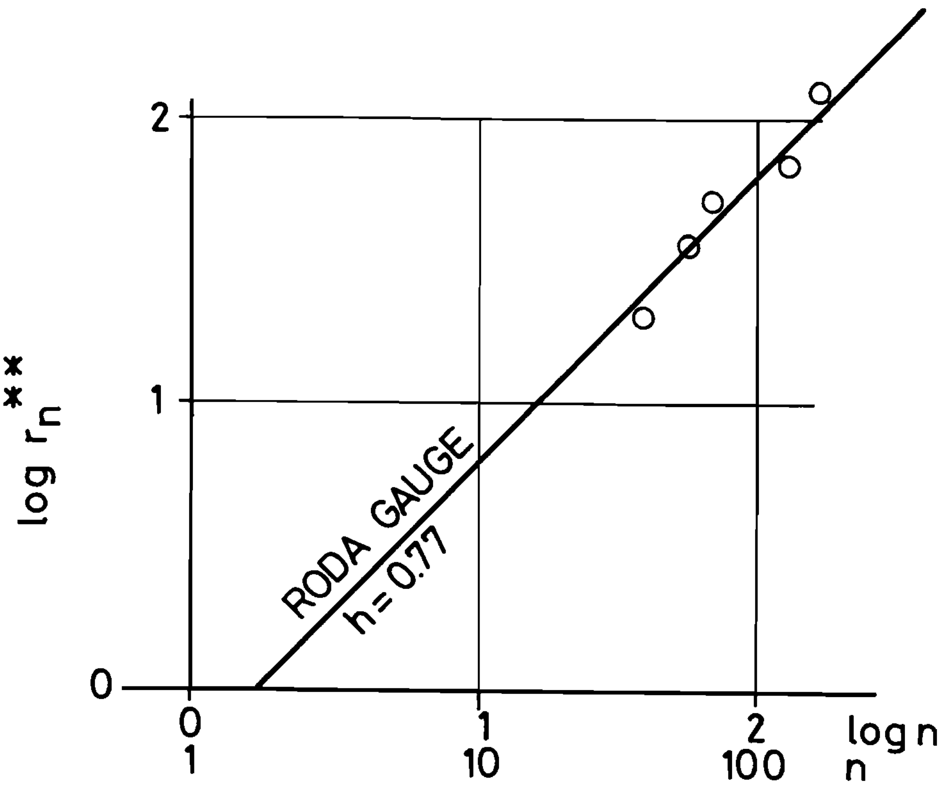

FIGURE 1

3.2 An Ambiguity

It is clear that Hurst's method was to plot values of $\log r_n^{**}$ against $\log n$, for a variety of values of n, for each of his sets of river data. In all cases the points appeared to lie on a straight line of slope h (where h is about 0.7), and pass through or near the point ($r_n^{**} = 1$, $n = 2^{-h}$).

What is not clear is the relation of a graph of this kind, involving perhaps five or six plotted points, to a single run of data in the summary Table 1. In our interpretation, we have assumed that the individual annual flow $x_1, x_2, \ldots, x_n$ corresponding to a typical run in the summary table here have been broken up into segments

$$\begin{aligned} & x_1, x_2, \ldots, x_{n(1)} \quad , \\ & x_{n(1)+1}, \ldots, x_{n(2)} \quad , \\ & \quad \cdot \\ & \quad \cdot \\ & \quad \cdot \\ & x_{n(s)+1}, x_{n(s)+2}, \ldots, x_n \quad , \end{aligned} \tag{7}$$

of convenient lengths, and then analyzed as follows.

1) For the first segment, with n(1) entries, compute the sample mean $\bar{x}_{n(1)}$, the accumulated deviations $x_1 + \ldots + x_r - r\bar{x}_{n(1)}$, $r = 1, 2, \ldots, n(1)$, the $\max(m_{n(1)})$ and the $\min(\ell_{n(1)})$ of these deviations, the adjusted range $r^*_{n(1)} = m_{n(1)} - \ell_{n(1)}$, where

$$d_n^2 = \sum_{j=1}^{n(1)} \{x_j - \bar{x}_{n(1)}\}^2 / n(1) \quad ,$$

and the rescaled adjusted range

$$r^{**}_{n(1)} = r^{*}_{n(1)}/d_{n(1)} \quad .$$

This value of $r^{**}_{n(1)}$ is then plotted against n(1), giving a single point on the log-log graph.

2) This first segment

$$x_1, x_2, \ldots, x_{n(1)}$$

is then enlarged by including the flows

$$x_{n(1)+1}, x_{n(1)+2}, \ldots, x_{n(2)}$$

which form the second row of (7) thus creating an extended segment

$$x_1, x_2, \ldots, x_{n(1)}, x_{n(1)+1}, \ldots, x_{n(2)} \quad ,$$

containing n(2) flow values (which include the n(1) flow values considered in the first segment). The extended segment is then analyzed in exactly the same way as was the first segment; that is we compute the sample mean $\bar{x}_{n(2)}$ of the extended segment and consider the appropriate accumulated deviations

$$x_1 + x_2 + \cdots + x_r = r\bar{x}_{(2)} \quad ,$$

for

$$r = 1, 2, \ldots, n(2) \quad .$$

We compute <u>their</u> max $m_{n(2)}$ and min $\ell_{n(2)}$, <u>their</u> standard deviation $d_{n(2)}$, where

$$d^2_{n(2)} = \sum_1^{n^2} \{x_j - \bar{x}_{n(2)}\}^2/n(2) \quad ,$$

and <u>their</u> rescaled adjusted range

$$r^{**}_{n(2)} = r^{*}_{n(2)}/d_{n(2)} \quad .$$

This computation provides a second point to be plotted at $\{n(2), r^{**}_{n(2)}\}$ on the log-log graph.

3) The enlarged segment is now further enlarged by adjoining the third segment, to become

$$x_1, x_2, \ldots, x_{n(2)} x_{n(2)+1}, \ldots x_{n(3)} \quad ,$$

and this is treated in the same way, etc. In this way one obtains a number of points,

$$\{n(1), r^{**}_{n(1)}\} \quad , \quad \{n(2), r^{**}_{n(2)}\} \quad , \quad \{n(3), r^{**}_{n(3)}\} \quad , \quad \text{etc.}$$

(where

$$n(1) < n(2) < n(3) \quad , \quad \text{etc.})$$

perhaps five or six in number, to which a reasonable looking straight line may be fitted in log-log graph paper.

(Our model, to be described in a subsequent section, would also be consistent with an alternative interpretation, namely that each of the "segments" corresponding to the rows of Table 1 has been separately analyzed by method 1) above, that is to say by working with non-overlapping segments, for each of which its own mean, its own max and min of accumulated deviations, its own standard deviation, etc. is computed. However, this does not seem to be the method that Hurst actually employed. There is another possible interpretation of Hurst's arithmetic, which is this. One computes the adjusted ranges $r^{*}_{n(1)}, r^{*}_{n(2)}, \ldots,$ of this first segment of n(1) items, the augmented segment counting of the first n(2) items, and so on, as described above, but rescales them all by dividing by a common divisor d_n, the standard deviation of the entire record of n flows. The relation of this interpretation to our model is discussed at the end of Section 5.)

3.3 Another Ambiguity

As we have explained, Hurst plotted his computed rescaled adjusted ranges r_n^{**} against n on log-log paper. The published graphs leave no doubt that a straight line plot is appropriate. The natural procedure would be to take the equation of the line as

$$\log r_n^{**} = \text{long } c + h \log n \quad , \qquad (8)$$

corresponding to the exponential curve

$$r_n^{**} = cn^h \quad . \qquad (9)$$

Here c, as well as h, would be estimated from the data. Hurst, however, appears to have convinced himself that the value of c ought to be taken as 2^{-h} , (i.e. log c = h log 2), so that in his formulation (8) became

$$\log r_n^{**} = h \log (n/2) \quad , \qquad (10)$$

and (9)

$$r_n^{**} = (n/2)^h \quad . \qquad (11)$$

Even if c is in fact near to 2^{-h} this procedure introduces a risk of producing a biased estimate of h, as exemplified in Figure 2.

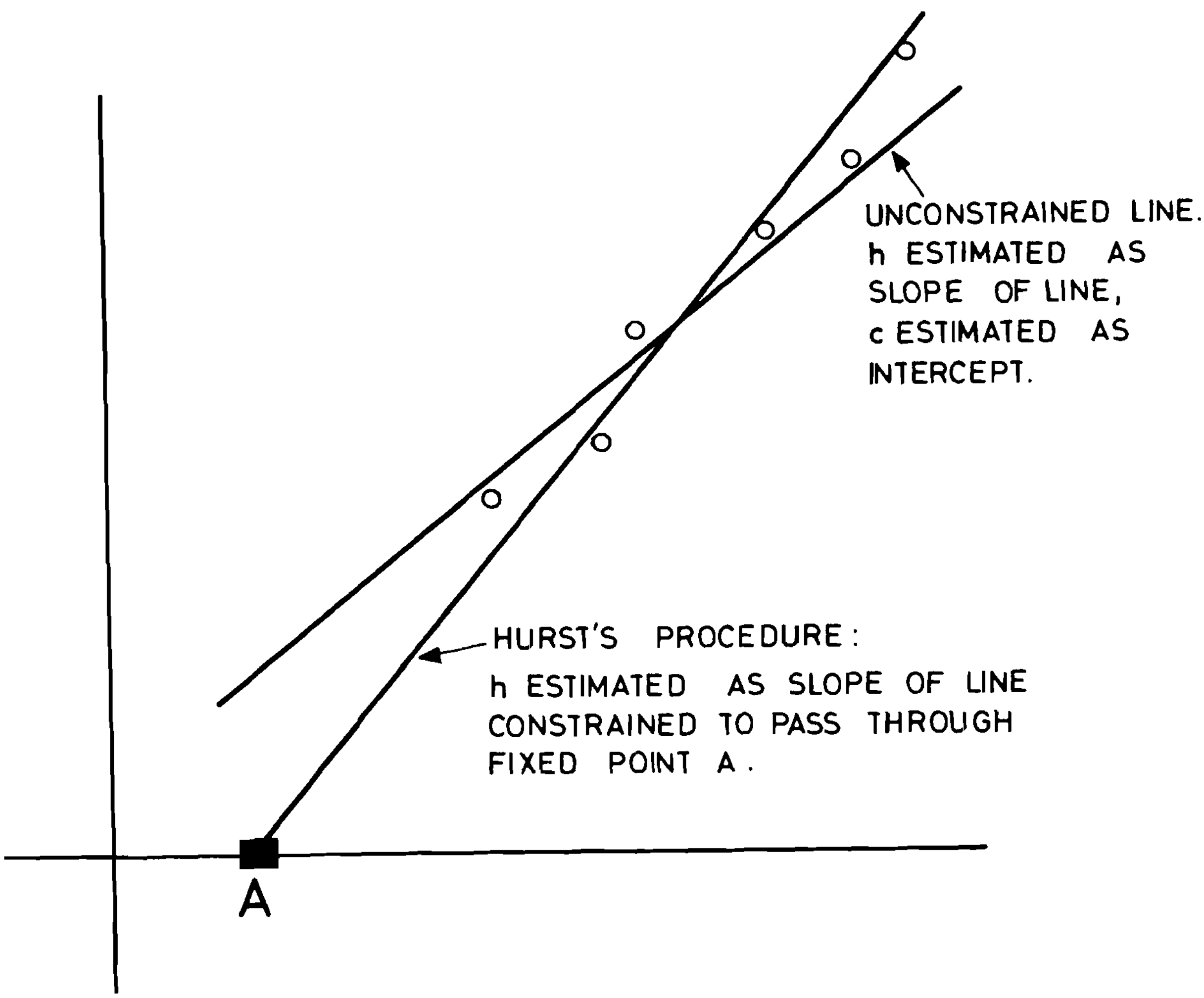
UNCONSTRAINED LINE.
h ESTIMATED AS
SLOPE OF LINE,
c ESTIMATED AS
INTERCEPT.
HURST'S PROCEDURE:
h ESTIMATED AS SLOPE OF LINE
CONSTRAINED TO PASS THROUGH
FIXED POINT A.
A

FIGURE 2

4. Hurst's Law: A Reformulation of the Empirical Version, and a Proposed Theoretical Model

In our view it is necessary to restrict the values of n for which validity is claimed for Hurst's law: the data on which the relationship rests belong to intervals of duration not less than about thirty years, and (with one or two doubtful exceptions) not more than about 1,000 years. There does not seem to be any scientific justification for assuming that the same rather simple relation should continue to hold for substantially longer intervals. The appropriate expression of Hurst's empirical law is therefore

$$r_n^{**} \propto n^h \text{ (ca. } 30 < n < \text{ ca. } 1{,}000) \quad , \qquad (12)$$

where h is approximately equal to 0.72. With the added phrase defining the relevant values of n (12) is a version of (9). We prefer to express r_n^{**} as being proportional to n^h, since we regard the essence of Hurst's discovery as lying in the value of h, the value of the constant c of (9) being of minor importance.

Before turning to an attempted formulation of a mathematical model it might be as well to note that such a model might reasonably reproduce the Hurst effect without having any pretensions to reflecting the "real" structure of the geophysical stochastic processes involved. Indeed it is far from clear what the "real" underlying structure is. For example, is it stationary? Is such a question meaningful? In the authors' view the mathematical concept of stationarity is a convenient simplification which is probably applicable, as a fair approximation, over a limited period, e.g. over a period of the same order of magnitude as the duration of the historical record; one would certainly not be justified without further evidence in postulating a continuation of stationarity into the indefinite future.

If this view is accepted, it would lead us to seek models for our data generation which would be valid for predictions over a period of the order of some hundreds of years but not necessarily beyond that time. Such models ought to reproduce the Hurst effect, but we do not believe that we would be justified in requiring of them that they continue to reproduce the Hurst effect for values of n of the order of 10^4 or larger, since, as we have already emphasized, the historical evidence does not necessarily imply such a time horizon.

Thus, in contrast with some of our distinguished colleagues (including Mandelbrot and his co-workers) who have interpreted the Hurst effect as implying the existence of an extremely long-term persistence in geophysical data, we have been concerned rather with the investigation of relatively simple models which might display Hurst-like behaviour over a period of up to about 1,000 years. If for such models the Hurst effect as we know it ceased to be manifested, or showed itself only in an attenuated form, for time intervals exceeding 1,000 years, this fact would not in our view invalidate the model.

Our aim has been to investigate analytically the statistical properties of the rescaled adjusted range of identically distributed random variables. We would of course like them to be autocorrelated, but so far (with one exception) we have been successful only with mutually independent variables. This work will be outlined in Section 5.

For reasons of space we must allow ourselves only the briefest mention of the large body of numerical work carried out in the field of simulation by various researchers, including Yevjevich (working with seasonally varying autoregressive processes), Mandelbrot and his co-workers (using "fractional Gaussian noise"), O'Connell (using "ARIMA" models involving a combination of autoregression and moving average) and Klemes (using a variety of distributions both independent and autocorrelated).

5. Our Interpretation of Hurst's Results

All of Hurst't data exhibited variability (which data do not?). For each river we may regard the data as a sample from a population of values. It is therefore appropriate to regard Hurst's values of r_n^{**} as the observed values of a random variable R_n^{**}. We define this random variable as follows: Let $X_1 X_2, \ldots, X_n$ represents n identically distributed random variables (the consecutive annual flows), and let $\bar{X}_n = \sum_{j=1}^{n} X_j/n$. Let

$$M_n = \max \{X_1 - \bar{X}_n, X_1 + X_2 - 2\bar{X}_n, \ldots, X_1 + \cdots + X_{n-1} - (n-1)\bar{X}_n, 0\} \qquad (13)$$

and

$$L_n = \min \{X_1 - \bar{X}_n, X_1 + X_2 - 2\bar{X}_n, \ldots, X_1 + \cdots + X_{n-1} - (n-1)\bar{X}_n, 0\} \qquad (14)$$

Then

$$R_n^{**} = (M_n - L_n)/D_n \quad ,$$

where

$$D_n^2 = \sum_{j=1}^{n} (X_j - \bar{X}_n)^2/n \quad .$$

We interpret Hurst's law in the empirical form (12) as meaning

$$\left.\begin{array}{l} E(R_n^{**}) \;\alpha\; n^h \\ \text{or} \\ E\left(\dfrac{M_n - L_n}{D_n}\right) \;\alpha\; n^h \end{array}\right\} \quad \text{ca. } 30 < n < \text{ca. } 1{,}000 \quad , \qquad (15)$$

(where "E(•)" denotes "expectation" in the statistical sense).

It is important to recognize that this is not the same as

$$\frac{E(M_n - L_n)}{E(D_n)} \; \alpha \; n^h \quad , \tag{16}$$

or

$$E(M_n - L_n) \; \alpha \; n^h \quad . \tag{17}$$

The fact that Hurst used the symbol "σ" to represent the scaling divisor D_n should not mislead us into regarding this as a known constant: it is an observed value d_n of a random variable D_n, and is subject to sampling variability in exactly the same way as is the numerator $M_n - L_n$ in (15).

The importance of this point can hardly be overemphasized. In attempting to build a theoretical model for the Hurst effect we would postulate some distributional form for the set $(X_1, X_2, \ldots, X_n)$, and then examine the probability distribution of the random variable $(M_n - L_n)/D_n$. It might be said at this point that investigations of this kind have not gone very far, but at least something is known about the expected value $E\{(m_n - L_n)/D_n\} = E(R_n^{**})$.

Even for the expectations, however, results have only recently become available (Anis and Lloyd [3]). Earlier workers, including ourselves, either found this random variable to be intractable or failed to appreciate the role of Hurst's scaling procedure. In the next section we shall outline some of this earlier work, which is largely restricted to evaluation of the expectation of the unscaled adjusted range $R_n^* = M_n - L_n$, standardized by division by the assume population value σ of the standard deviation of the inflow X_j (which we may conveniently take to be unity) or even of the unscaled (and unadjusted) crude range R_n, similarly standardized, where

$$R_n = \max(0, X_1 - \mu, X_1 + X_2 - 2\mu, X_1 + \cdots + X_n - n\mu)$$

$$-\min(0, X_1 - \mu, X_1 + X_2 - 2\mu, X_1 + \cdots + X_n - n\mu) \quad ,$$

the population value σ of the common expectation of the common expectation of the X_j replacing the sample mean $\bar{X}_n$ used in the adjusted range.

The real "justification" for using the unscaled adjusted range is that it is more amenable to mathematical treatment than is the true Hurst range. A more respectable but somewhat fallacious justification would be the argument that

$$E\left(\frac{M_n - L_n}{D_n}\right) \cong \frac{E(M_n - L_n)}{E(D_n)} \cong \frac{E(M_n - L_n)}{\sigma}$$

where σ denotes the population value of the inflow standard deviation. Since D_n is positively correlated with $M_n - L_n$ this approximation can lead to possibly substantial errors.

The effect of the positive correlation between the adjusted range $M_n - L_n$ and the sample standard deviation D_n is shown in an exaggerated form in the case where n = 2. In this case

$$\bar{X}_n = \bar{X}_2 = \tfrac{1}{2}(X_1 + X_2) \quad ,$$

$$X_1 = \bar{X}_2 = \tfrac{1}{2}(X_1 - X_2) = -(X_2 - \bar{X}_2)$$

and

$$nD_n^2 = 2D_2^2 = (X_1 - \bar{X}_2)^2 + (X_2 - \bar{X}_2)^2 = \tfrac{1}{2}(X_1 - X_2)^2 \quad ,$$

whence

$$D_2 = \tfrac{1}{2}|X_1 - X_2| \quad .$$

Thus

$$M_2 - L_2 = \max\{0,\tfrac{1}{2}(X_1 - X_2)\} - \min\{0,\tfrac{1}{2}(X_1 - X_2)\}$$

and

$$R_n^{**} = \max\{0,\text{sign }(X_1 - X_2)\} - \min\{0,\text{sign }(X_1 - X_2)\} \quad ,$$

since

$$\frac{X_1 - X_2}{|X_1 - X_2|} = \begin{Bmatrix} +1 & , & \text{if } X_1 > X_2 \\ -1 & , & \text{if } X_1 < X_2 \end{Bmatrix} = \text{sign }(X_1 - X_2) \quad .$$

(We may neglect the case $X_1 = X_2$ as having zero probability.) Considering the possible cases, we have

Case 1): $X_1 > X_2$, sign $(X_1 - X_2) = +1$,

$$R_2^{**} = \max(0,1) - \min(0,1) = 1 - 0$$
$$= 1 \quad .$$

Case 2): $X_1 < X_2$, sign $(X_1 - X_2) = -1$,

$$R_n^{**} = \max(0,-1) - \min(0,-1) = 0 - (-1)$$
$$= 1 \quad .$$

Thus in both cases the random variable R_2^{**} reduces to a <u>constant</u>, whatever the distribution of the X.

In contrast to this, the unrescaled adjusted range is

$$R_2^* = M_2 - L_2$$
$$= \max\{0, \tfrac{1}{2}(X_1 - X_2)\} - \min\{0, \tfrac{1}{2}(X_1 - X_2)\}$$
$$= \tfrac{1}{2}|X_1 - X_2| \quad ,$$

a random variable whose expectation is necessarily sensitive to the distribution of the X_j.

(The "other possible interpretation" of Hurst's work mentioned at the end of Section 3.2 does not lend itself to a clear theoretical formulation, producing a "range" of the form

$$\{M_{n(r)} - L_{n(r)}\}/D_n \quad , \qquad r = 1,2,\ldots,s \quad ,$$

where

$$n(1) < n(2) < \cdots \leq n \quad ,$$

and where the process relation of the n(r) to n would have to be taken into account. Perhaps the best way to deal with this would be to regard it as an intermediate case between our (14) and a new range

$$R^{***}_{n,m} = (M_n - L_n)/D_m$$

where $M_n - L_n$ is defined as in (14) and

$$D^2_m = \sum_{j=1}^{m} \{X'_j - \bar{X}'_m\}^2/m \quad ,$$

the variables $X'_1, X'_2, \ldots, X'_m$ representing a set of m flows which are independent of the n flows and used in defining $M_n - L_n$. This is a well-defined random variable, whose properties, however, have not been investigated.)

6. A Brief Summary of Stochastic Models

Sums of independent random varibles have long been objects of interest to probabilists, and it is well-known that, for the crude range R_n of sums of independent and identically distributed random variables having finite variance, the expectation satisfies

$$E(R_n) \sim n^{1/2}$$

for sufficiently large n. An approximate value of $E(R^*_n)$ for binomial increments was obtained by Hurst.

The exact value of $E(R_n)$ for finite values of n with independent Normal X's obtained by Anis and Lloyd. Subsequent investigations of the unscaled adjusted range R^*_n (for sums of independent increments X_i) were carried out by Feller (Brownian motion) [10], Solari and Anis (Normal increments)[17], Moran ("stable" increments-crude range only)[16], Boes and Salas-La Curz ("stable")[7] and other increments, adjusted range), Moran (gamma distributed-crude range only)[16] and Anis and Lloyd (gamma-adjusted range)[2].

Recently the expectation of the rescaled adjusted range (15) has been obtained for the case of independent normal increments and also for a special case of increments having a multivariate, normally correlated distribution. The results of these investigations may be summarized as in Table 2. The only known theoretical basis for comparison between analytical results on the crude, the adjusted, and the Hurst range rests on the cases of independent normal inflows.

The formulae given in Table 2 yield functions whose graphs are qualitatively of the following form:

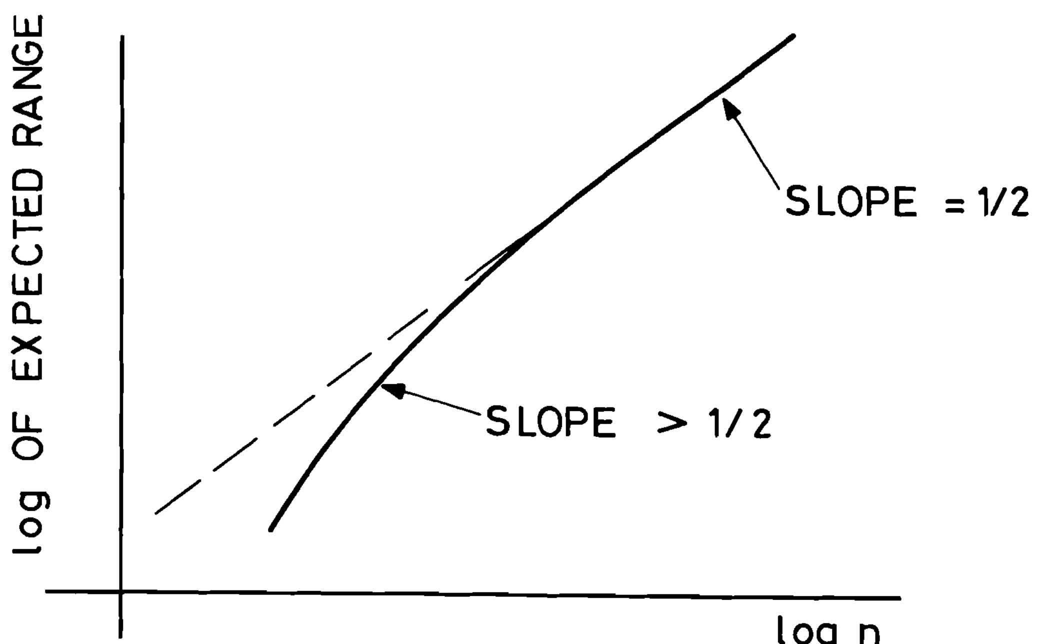

FIGURE 3

Table 2.

INDEP		CRUDE	ADJUSTED	RESCALED & ADJUSTED	
Inflow distribn	Author(s)	$E(R_n)$	$E(R_n^*)$	$E(R_n^{**})$	Comments
Normal	Feller	$n^{\frac{1}{2}}$	$n^{\frac{1}{2}}$		Valid for large n only
Normal	Anis/Lloyd	$\sum_1^n r^{-\frac{1}{2}}$			(Asympt. like $n^{\frac{1}{2}}$)
Normal	Solari/Anis		$n^{\frac{1}{2}} \sum_1^{n-1} 1/\sqrt{s(r-s)}$		(Approx. equal to 2 $\sin^{-1}$ (1 - 1/n) Asympt.like $n^{\frac{1}{2}}$
Normal	Anis/Lloyd			$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma \frac{n}{2}} \sum_1^{n-1} \sqrt{\frac{n-1}{r}}$	
Gamma (m)	Anis/Lloyd		$\frac{\Gamma(mn)}{m^{\frac{1}{2}} n^{mn}} \sum_1^{n-1} \frac{k^{km-1}(n-k)^{(n-k)m}}{\Gamma(km)\Gamma\{(n-k)m\}}$		
Symm.	Kac/Pollard Darling				Asymptotic results only
Stable(α)	Moran	$\sum_1^n k^{\frac{1}{2}-1}$			(Asymp. like $n^{1/\alpha}$
	Boes and Salas		$\sum_1^n \left\{ r\left(\frac{1}{r} - \frac{1}{r}\right)^{\alpha} + \frac{n-r}{n^{\alpha}} \right\}^{\alpha}$		
CORR					
NORMAL AUTOREG etc.	Yevjevich	$\sum_1^n \frac{\mathrm{var}(X_1 + \cdots + X_r)}{r}$			A conjectured formula. shown by simulation to be a good approx.
SYMM MULTINORL	Boes/Salas		$\frac{1}{\sqrt{n}} \sqrt{1-\rho} \sum_1^{n-1} \sqrt{\frac{n-r}{r}}$		
SYMM MULTINORL	Anis/Lloyd			$\frac{\Gamma\left(\frac{n-r}{2}\right)}{\Gamma \frac{n}{2}} \sum_1^{n-1} \sqrt{\frac{n-r}{r}}$	Independent of Correlation Coefficient ρ

Quantitative results are given in Table 3, in which we give also the corresponding results for the unrescaled adjusted range $E(R_n^*)$. The slope at a given value of n may be regarded as the "<u>local Hurst exponent</u>" h(n).

Table 3. Adjusted range $E(R_n^{**})$, with corresponding values of local Hurst exponent h(n), for independent Normal increments.

n	Crude range $E(R_n)$	h(n)	Adjusted range $E(R^*_n)$	h(n)	Rescaled range $E(R_n^{**})$	h(n)
5	2.58	.67	1.62	.89	1.93	.68
10	4.01	.61	2.79	.71	3.02	.63
20	6.06	.58	4.44	.63	4.61	.59
50	10.17	.55	7.70	.58	7.81	.56
100	14.83	.54	11.39	.55	11.45	.54
n-72		.5		.5		.5

It will be seen that the unscaled adjusted range overestimates the local Hurst exponent. If it were established that the relation between the unscaled adjusted range and the Hurst range were similar to this for other distributional forms such as the gamma (and by continuity arguments this must be so for gamma distributions of small skewness) the results obtained for the adjusted range of gamma and stable inflows, tabulated in Table 2, could be regarded as relevant to our discussion. For these the local Hurst exponents for the <u>rescaled</u> adjusted range, without being too poor an approximation to the latter) are as follows in Table 4.

Table 4. Value of the local Hurst exponent h(n) for the unrescaled adjusted range of gamma (m) inflows.

n	h(n)		
(length of record)	m = .100	m = .010	m = .001
10	.89	1.07	1.11
20	.76	.97	1.04
50	.65	.86	1.00
100	.61	.78	.96
500	.54	.63	.84
1,000	.53	.59	.77

It will be seen by interrolation that mutually independent gamma (m) inflows, with m ≅ 0.005, reproduce the Hurst effect very well in the desired interval of say, fifty to 1,000 years, and it is reasonable to suppose that similar results would hold for the Hurst range of independent gamma inflows having a shape parameter m not very different from 0.005.

It must be admitted that this shape parameter represents an unrealistically high degree of skewness. It is possible that similar results might be obtained with a more acceptable skewness in terms of a more flexible inflow distribution family. Work is proceeding along these lines for the log-Normal and the non-central chi-squared families.

7. The Effect of Correlation

It would of course be completely unrealistic to pretend that the annual increments X_i are in fact mutually independent. The independence assumption implied in the results of Section 5 has been forced upon us by reasons of tractability:

that is, workers who have assumed independence have done so in the hope of developing methods which are capable of being generalized to deal with correlation. So far it must be admitted that only in one case has theory proved capable of dealing with correlated flows, namely the situation when the Normal increments X_i are "symmetrically correlated," that is where for example corr (x_1,x_2) = corr $(x_1,x_3) = \cdots =$ = corr (x_1,x_n) = corr $(x_2,x_3) = \cdots =$ corr $(x_{n-1},x_n) = \rho$.

No conceivable geophysical system could behave in this way. The results obtained are nevertheless not without interest, since it turns out that for the unrescaled adjusted range the expectation is proportional to $(1 - \rho)^{\frac{1}{2}}$, whereas in the rescaled adjusted case (the Hurst range) the expectation does not depend on ρ at all.

That such an unexpected result could exist, albeit in an unrealistically correlated situation, must point to a need for caution in extrapolating from the unrescaled results to the Hurst range.

Acknowledgments

The authors wish to express their appreciation of the hospitality offered by IIASA during the month of May 1975 when this work was carried out.

References

[1] Anis, A.A., and Lloyd, E.H. "On the Range of Partial Sums of a Finite Number of Independent Normal Variates." Biometrika, 40 (1953), 35-42.

[2] Anis, A.A., and Lloyd, E.H. "Skew Inputs and the Hurst Effect." Int. J. Hydrol, 26 (1975), 39-54.

[3] Anis, A.A., and Lloyd, E.H. "The Expected Value of the Rescaled Hurst Range of Independent Normal Summands." (Forthcoming).

[4] Bachelier, L. "Les Lois des Grands Nombres du Calcul de Probabilites." Paris, Gauthier-Villars, 1937.

[5] Baxter, G., and Donsker, M.D. "On the Distribution of the Supremum Functional of Processes with Stationary Independent Increments." Trans. Amer. Math. Soc., 85 (1957), 73-87.

[6] Berman, S.M. "Limiting Distributions of the Maximum of a Diffusion Process." Annals of Math. Stat., 22 (1951), 427-432.

[7] Boes, D.C., and Salas-La Cruz, J.D. "On the Expected Adjusted Range of Partial Sums of Exchangeable Random Variables." J. Appl. Prob., 10 (1973), 671-677.

[8] Darling, D.A. "The Maximum of Sums of Stable Random Variables." Trans. Amer. Math. Soc., 83 (1956), 164-169.

[9] Erdoes, P., and Kac, M. "On Certain Limit Theorems of the Theory of Probability." Bull. Amer. Math. Soc., 52 (1946), 292-302.

[10] Feller, W. "The Asymptotic Distributions of the Range of Independent Random Variables." Annals of Math. Stat., 22 (1951), 427-432.

[11] Hurst, H.E. "Long-Term Storage Capacity of Reservoirs." Trans. Amer. Soc. Civ. Engin., 116 (1951), 770-799.

[12] Hurst, H.E., Black, R.P., and Simnika, Y.M. Long-Term Storage. London, Constable, 1965.

[13] Kac, M., and Pollard, H. "The Distribution of the Maximum of Partial Sums of Independent Random Variables." Can. J. Maths., 2 (1950), 375-384.

[14] Klemes, V. "The Hurst Phenomenon: A Puzzle?" Water Resources Res., 10 (1974), 674-688.

[15] Levy, P. Processus Stochastiques et Mouvement Brownien. Paris, Gauthier-Villars, 1948.

[16] Mandelbrot, B. "Une Classe de Processus Stochastiques Homothetiques a Soi; Application a la Loi Climatologique de H.E. Hurst." C.R. Acad. Sc. Paris, 260 (1965), 3274-3277.

[17] Moran, P.A.P. "On the Range of Cumulative Sums." Ann. Inst. Maths. Toyko, 16 (1964), 109-112.

[18] Solari, M.E., and Anis, A.A. "The Mean and Variance of the Maximum of the Adjusted Partial Sums of a Finite Number of Independent Normal Variates." Annals of Math. Stat., 28 (1957), 706-716.

[19] Yevjevich, V. "Mean Range of Linearly Dependent Normal Variables with Application to Storage Problems." Water Res. Research, 3 (1967), 663-671.