

Computer Experiments with Fractional Gaussian Noises. Part 1, Averages and Variances

BENOIT B. MANDELBROT

AND

JAMES R. WALLIS

*International Business Machines Research Center
Yorktown Heights, New York 10598*

Abstract. Fractional Gaussian noises are a family of random processes such that the interdependence between values of the process at instants of time very distant from each other is small but nonnegligible. It has been shown by mathematical analysis that such interdependence has precisely the intensity required for a good mathematical model of long run hydrological and geophysical records. This analysis will now be illustrated, extended, and made practically usable, with the help of computer simulations. In this Part, we shall stress the shape of the sample functions and the relations between past and future averages.

Using a computer and a graphic plotter as tools, we have carried out simulation experiments concerning the sample behavior of 'fractional Brownian motions' and of 'fractional Gaussian noise.' Most of the results thus far obtained will be described in this paper. Earlier [Mandelbrot, 1965; Mandelbrot and Wallis, 1968a] fractional noises were presented as models of hydrological records. A later paper provides empirical evidence for this argument [Mandelbrot and Wallis, 1969a].

Fractional Brownian motions and fractional Gaussian noises, as defined in Mandelbrot and Van Ness [1968], are, respectively, generalizations of 'Brownian motion' and of 'white Gaussian noise.' The latter term is a synonym for 'sequence of independent Gauss random variables.'

The term 'fractional noise' can be explained by considerations from spectral theory. (The bulk of this paper can be followed even if this explanation is not fully appreciated.) Classically, a white noise is defined as a random process having a spectral density independent of the frequency f . As a result, the integral and derivative of white noise, and its repeated integrals or derivatives, all have spectral densities of the form f^{-2k} , with k as an integer. Fractional noises, on the contrary, can be defined as having

a spectral density of the form f^{-2k} , with k a non-integer fraction. This explains the term 'fractional noise,' or more precisely 'fractionally white noise.' By repeated integration or differentiation, one can restrict k to any interval of unit length. For reasons that will become apparent, k will be written as $-H + 0.5$, where H , the main parameter of a fractional noise, will be left to vary between 0 and 1. Classical white noise is yielded by $H = 0.5$. Despite their role in the above definition, spectral techniques will be the least important among the many used in this paper.

For a list of notations, see p. 266 at the end of Part 3.

Two approximations to fractional Gaussian noise were employed in our simulations. Our 'Type 2 approximation' is grosser and less important but easier to define. It is given by the two-parameter moving average

$$F_2(t | H, M) = (H - 0.5)$$

$$\sum_{u=t-M}^{t-1} (t-u)^{H-1.5} G(u) + Q_H G(t)$$

In this definition, $G(u)$ is a sequence of independent Gauss random variables of zero mean and unit variance. The constant Q_H depends upon H , as follows:

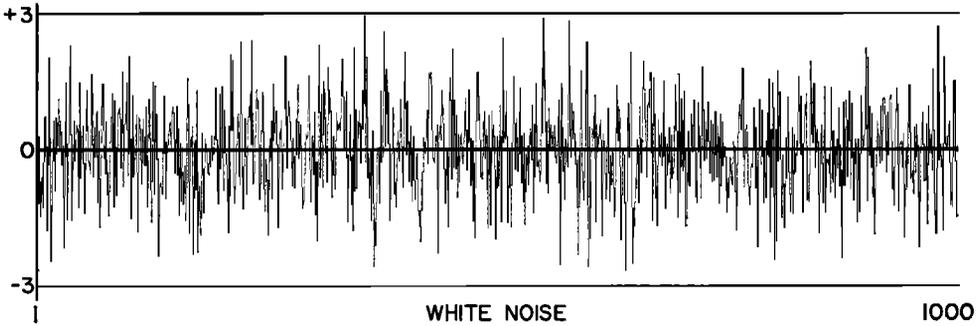


Fig. 1. A sample of 1000 values of discrete white noise of zero mean and unit variance, also called sequence of independent Gauss random variables, also called discrete fractional noise with $H = 0.5$.

$$Q_H = 0 \quad \text{if} \quad 0.5 < H < 1,$$

and

$$Q_H = (0.5 - H) \sum_{u=1}^{\infty} u^{H-1.5} \quad \text{if} \quad 0 < H < 0.5$$

The final parameter M , called the ‘memory of the process,’ is some very large quantity; originally [Mandelbrot, 1965], M was set to $M = \infty$, but in the experiments to be reported, M was varied from 1 to 20,000.

The definitions of fractional Gaussian noise itself and of ‘Type 1 approximation’ to it are more cumbersome, and are postponed to Part 3, the Mathematical Appendix.

For actual records [Mandelbrot and Wallis, 1969a] the Gaussian assumption is only an approximation, and the nonGaussian fractional noises [Mandelbrot and Van Ness, 1968] are very important. References to them will, however, be postponed; therefore, the epithet ‘Gaussian’ can be omitted in the present paper, and we shall simply speak of ‘fractional noises.’

THE SHAPE OF THE SAMPLE FUNCTIONS OF FRACTIONAL NOISE

We begin by exhibiting on Figure 1 a sample of a process of independent Gauss random variables. A short sample suffices, because this process is monotonous and ‘featureless.’ It is analogous to the ‘hum’ in electronic amplifiers and is often called a ‘discrete-time white noise.’ It can also be considered a discrete-time fractional noise with $H = 0.5$.

Another small sample, with $H = 0.1$, is given as Figure 2. Visually, it does not differ much from the white noise $H = 0.5$, but statistical tests show it to be richer in ‘high frequency’ terms, owing to the fact that large positive values ‘tend to’ be followed by ‘compensating’ large negative values.

Friezes 1 to 8 carry successive samples, each containing 1000 values of a moderately nonwhite fractional noise with $H = 0.7$. Similarly, Friezes 9 to 16 carry a strongly nonwhite fractional noise, with $H = 0.9$. Whenever $H > 0.5$, a fractional noise is richer than

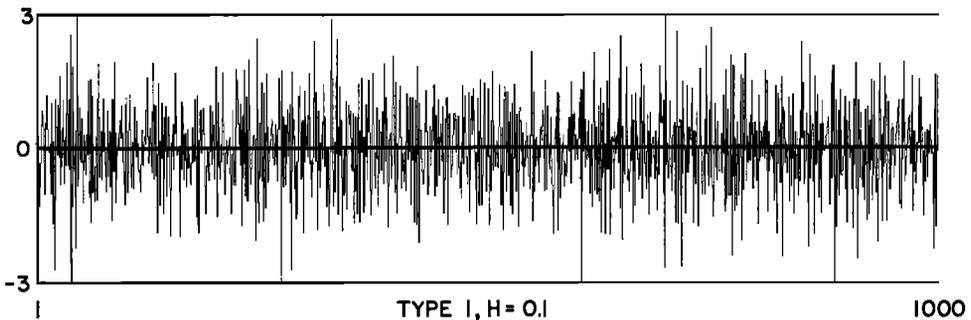
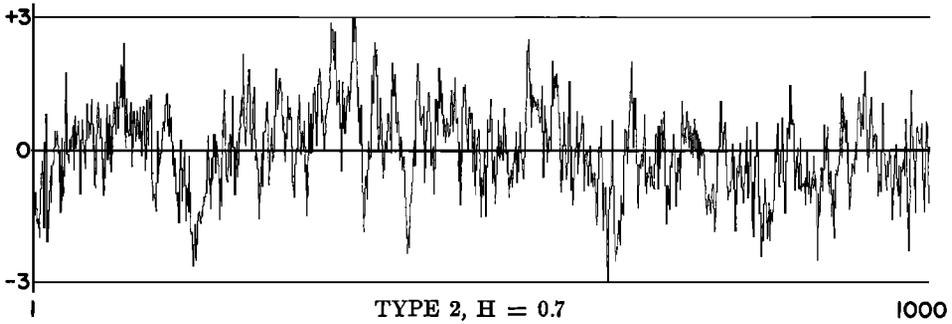


Fig. 2. A sample of 1000 values of a Type 1 approximate fractional noise with $H = 0.1$. The process was normalized to have zero mean and unit variance.



Frieze 1. The first one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$. The process was normalized to have zero mean and unit variance. The type 2 approximation was selected because a Type 1 with $H = 0.7$ has so much high frequency 'jitter' that low frequency effects are illegible.

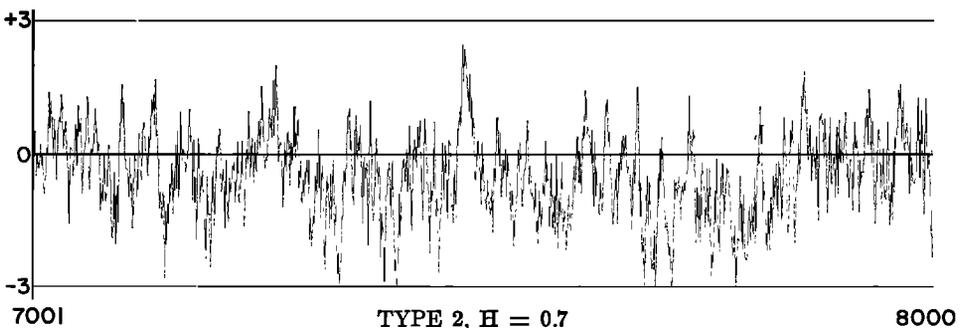
white noise in 'low frequency' terms, owing to the fact that large positive or negative values tend to 'persist.' Even a casual glance at these figures shows the effect of such low frequency terms. We wish to encourage the reader to compare our artificial series with the natural records with which he is concerned. To be meaningful, the comparison must involve in both cases the same degree of local smoothing of high frequency 'jitter.' If the artificial series 'feel' very different from his natural records, then fractional noises are probably inappropriate for his problem. If they feel close, but not quite right, then perhaps another value of H will suffice. If they feel right, then he should proceed to further and more formal statistical tests of fit. Such formal tests are indispensable but should neither be blind nor come first. A statistical test by necessity focuses upon a specific aspect of a process, whereas the eye can often balance and integrate various aspects. We em-

phasize that formal test and visual inspection should be combined.

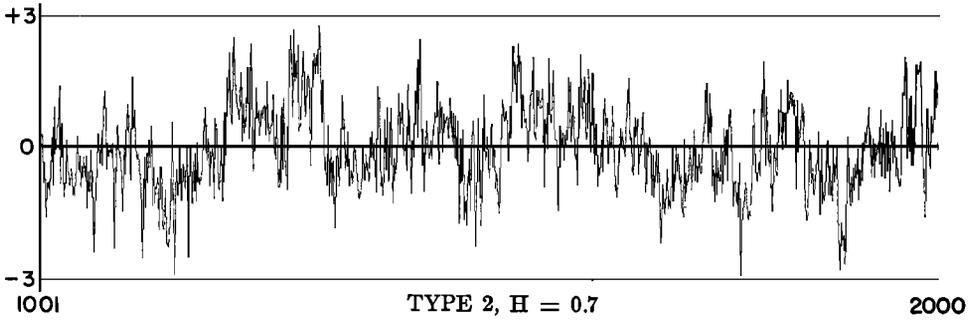
To cite our own experience, visual inspection helped us to establish that the H coefficient of hydrological records is not 'typically' near $H = 0.7$. Indeed, the sample noise with $H = 0.7$, which was the first to be plotted, seemed much too smooth and regular to be reasonable in all cases. The noise with $H = 0.9$ was plotted next, and it appeared that the 'typical' span of values of H should stretch at least from $H = 0.7$ to $H = 0.9$, other values such as $H = 0.5$ itself being possible. This conclusion has since been confirmed by direct test, as will be seen in *Mandelbrot and Wallis* [1969a].

FRACTIONAL NOISE WITH $0.5 < H < 1$, AND APPARENT 'CYCLES'

A perceptually striking characteristic of fractional noises is that their sample functions exhibit an astonishing wealth of 'features' of every



Frieze 8. The eighth one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.



Frieze 2. The second one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

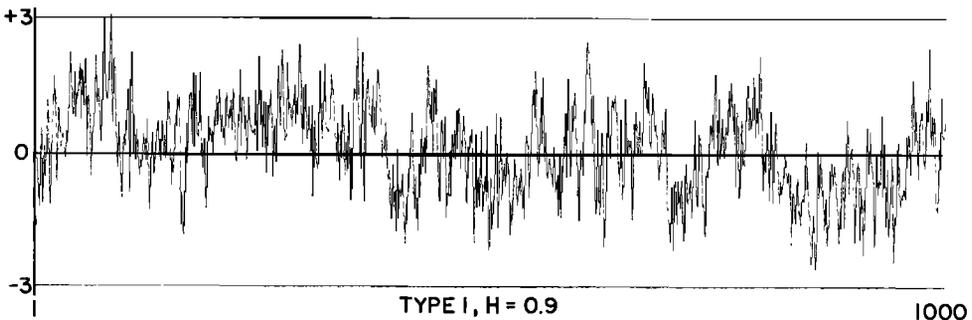
kind, including trends and cyclic swings of various frequencies. In some cases, such swings are rough and far from periodic, but some subsamples seem to include absolutely periodic swings. However, the wavelength of the longest among the apparent cycles markedly depends on the total sample size. If one looks at a shorter portion of these friezes, shorter cycles become visible. At the other extreme, for our original plots of these friezes as strips of 3000 time units, the impression was unavoidable that cycles of about 1000 time units were present. Since there is no built-in periodic structure whatsoever in the generating mechanism, such cycles must be considered spurious. But they are very real in the sense that something present in human perceptual mechanisms brings most observers to 'recognize' the same cyclic behavior. Such cycles are useful in describing the past but have no predictive value for the future. Analogous remarks were made by *John M. Keynes* [1939] about economic cycles. This is not surprising, since the resemblance between economic and

hydrological time series has been often noted. As a rough rule of thumb, unless the sample size is very short, the longest cycles have a wavelength equal to one-third of the sample size. This 'one-third rule' is not only a matter of psychology of perception, but also a consequence of the self-similarity of fractional noise, to be discussed in Part 3, the Mathematical Appendix.

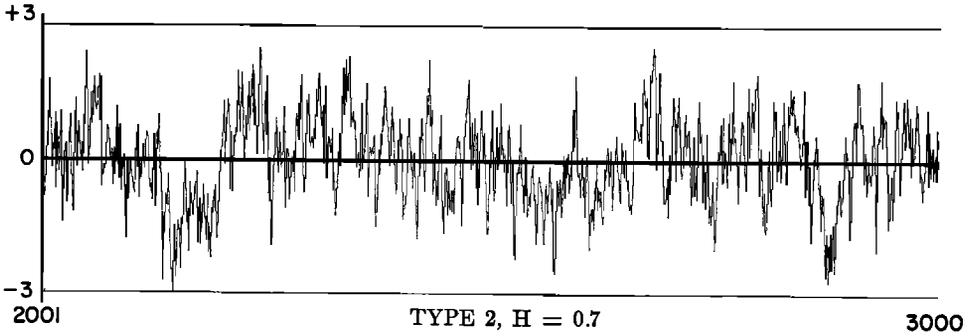
The classical mathematical technique for studying periodicities, namely spectral or Fourier analysis, will be examined later. Suffice it to say at this point that spectral analysis *does not* confirm the apparent periodic appearance of fractional noise.

FRACTIONAL NOISE AND 'RUNS'

A second and even more striking characteristic of fractional noise with $H > 0.5$ is that some periods above or below the theoretical mean, which equals 0 by construction, are extraordinarily long. In fact, portions of these figures are reminiscent of the seven fat and seven lean



Frieze 9. The first one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$. The process was normalized to have zero mean and unit variance. When $H = 0.9$, the high-frequency 'jitter' is not strong enough to cover the low frequency effects.



Frieze 3. The third one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

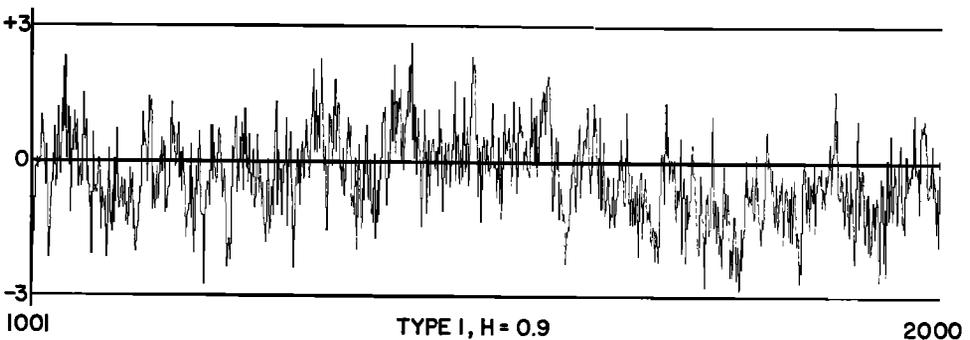
years of Joseph son of Jacob. One is tempted to express this perceptual 'persistence' of fractional noise with the help of the idea of a 'run.' A run of dry weather, or 'dry run,' could be defined as a period when precipitation stays below the line; a 'wet run' as a period when precipitation stays above the line. However, a careful inspection of fractional noise shows many instances where this concept of run describes its behavior very poorly. Often, one is tempted to call a period a wet run although it is interrupted by a very short dry run. Should we be pedantic and consider such a sequence as being three runs? Or is it 'really' a single run? Perhaps the short dry runs in question could be eliminated by a little smoothing? Such a chase after a reasonable definition of 'runs' had to be abandoned because we found it hopeless. We shall not clutter this paper with proofs of our failure. We tried scores of smoothing procedures (moving averages of various lengths) and scores of redefinitions of 'dry' and 'wet' (different crossover levels). The distribution of the dura-

tion of runs was found to depend very much on otherwise insignificant features of the model, whereas large differences in the relative proportions of low and high frequencies were obscured.

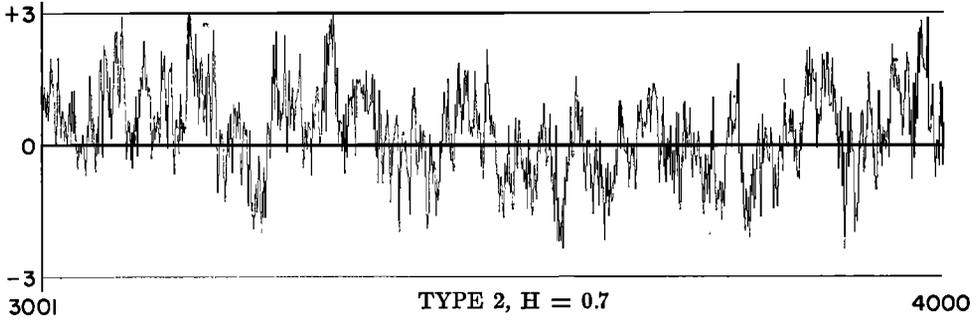
FRACTIONAL NOISE AND HIGH AND LOW 'TERMS'

Other authors have also found 'runs' to be too elusive to be useful. How to express, then, the perceptual evidence that fractional noise is 'persistent?' What we really see in such noise are fairly long 'terms' of years where, on the whole, the level is high, and other terms where it is low. The Biblical 'seven fat years' could be an example of a high term, the 'seven lean years' being an example of a low term. Mathematically, the average level during a high term may be exceeded only by a few extremes during the preceding or following low term. The group of sections that follows will suggest several ways of expressing precisely the concept of 'term.'

We shall find it necessary to use a formula proved in Part 3, the Mathematical Appendix. This formula applies to quantities of the form



Frieze 10. The second one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.



Frieze 4. The fourth one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

$$\Delta X^* = X^*(t + s) - X^*(t) = \sum_{u=1}^s X(t + u)$$

in case $X(t)$ is a discrete-time fractional noise and $X^*(t)$ is a fractional Brownian motion. In that case, $X^*(t)$ is to be denoted by $B_H(t) - B_H(0)$, and it will be shown that one has

$$\mathcal{E}[\Delta X^*]^2 = \mathcal{E}[\Delta B_H]^2 = C_H s^{2H}$$

where the sign \mathcal{E} denotes an expectation, and C_H is a positive constant.

In the case of approximations to fractional noise, the above formula only applies as an approximation valid for 'large' s , where the point at which s can be called 'large' depends upon H , and upon the approximation in question.

THE VARIANCE OF SECULAR AVERAGES

A first test of the perceptual impression that fractional noises with $H > 0.5$ are persistent is that averages over successive spans of s years differ markedly from the preset zero population expectation. We shall be concerned with values of s of the order of 50 or 100, and, following the

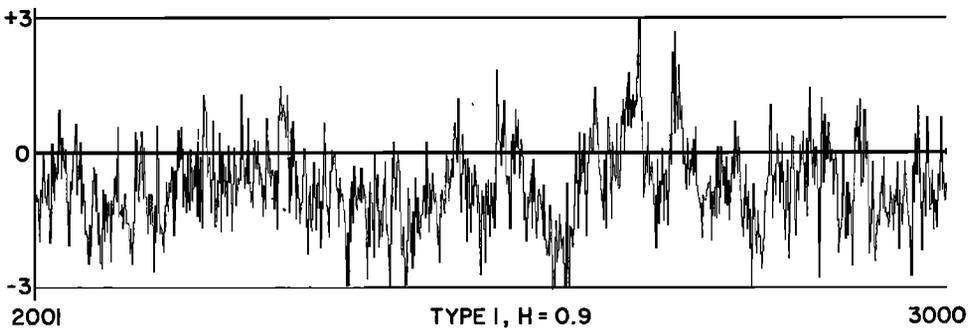
usage of astronomy, the corresponding averages will be called secular. Successive secular averages of a white noise, are, on the contrary, practically indistinguishable from one another. To account for this discrepancy of behavior, we shall now evaluate the variance of a secular average, about the origin, as a function of s and of H .

Fractional noises being Gaussian random processes, the sample average over s years is a Gaussian random variable. Its variance equals

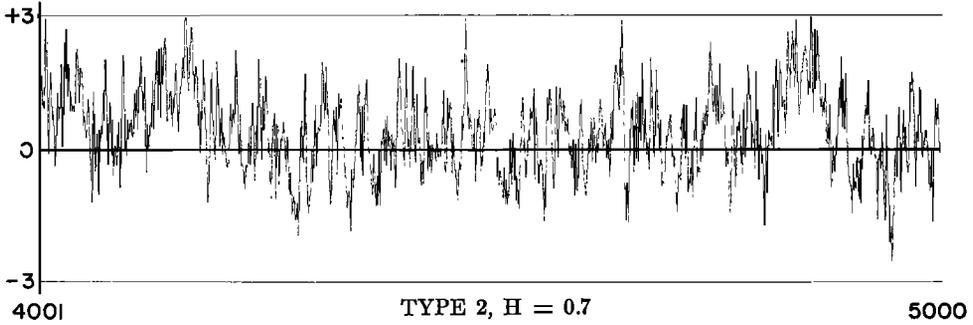
$$\mathcal{E} \left[s^{-1} \sum_{u=1}^s X(t + u) \right]^2 = C_H s^{2H-2}$$

so that its standard deviation is $\sqrt{C_H} s^{H-1}$. This expression always decreases to zero as $s \rightarrow \infty$, but the speed of its decrease depends sharply on H .

To give a numerical example, let $s = 100$. Then s^{H-1} equals 0.016 with $H = 0.1$, equals 0.10 with $H = 0.5$, equals 0.25 with $H = 0.7$, and equals 0.62 with $H = 0.9$. In other words, the ratio 'standard deviation of a hundred year average/standard deviation of a yearly reading' is still 2/3 when $H = 0.9$, it is only 1/4 when



Frieze 11. The third one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.



Frieze 5. The fifth one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

$H = 0.7$, it reduces to a mere 1/10 in the classical $H = 0.5$ case, and it becomes minute when $H = 0.1$.

When the standard deviation of a process decreases slowly by averaging, one is tempted to say that this process 'fluctuates violently.' With such a definition, we see that *fractional noises with a high value of H are the most violently fluctuating among fractional noises.*

Curves that represent the dependence upon s of the variance of the secular average are sometimes called 'variance time' curves by statisticians, and are credited by them to G. Udry Yule. They appear, in fact, to date back to G. I. Taylor's 1921 work on turbulent diffusion (see *Friedlander and Topper* [1961]).

THE SEQUENTIAL VARIANCE

We shall define $S^2(t, s)$ as the variance of s values $X(t + 1), \dots, X(t + s)$ around their sample average. Thus

$$S^2(t, s)$$

$$\begin{aligned} &= s^{-1} \sum_{u=1}^s \left[X(t + u) - s^{-1} \sum_{v=1}^s X(t + v) \right]^2 \\ &= s^{-1} \sum_{u=1}^s X^2(t + u) - \left[s^{-1} \sum_{u=1}^s X(t + u) \right]^2 \\ &= s^{-1} \sum_{u=1}^s X^2(t + u) - [s^{-1} \Delta X^*]^2 \end{aligned}$$

Consequently

$$\mathcal{E}S^2(t, s) = X^2 - \mathcal{E}[s^{-1} \Delta X^*]^2$$

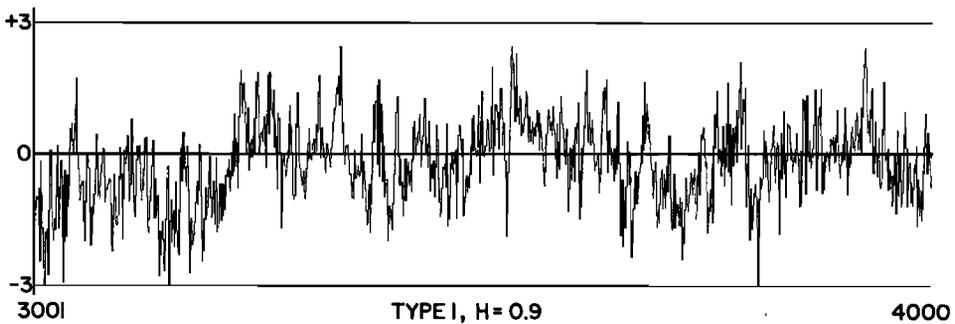
For discrete time fractional noise, this becomes

$$\mathcal{E}S^2(t, s) = C_H - C_H s^{2H-2}$$

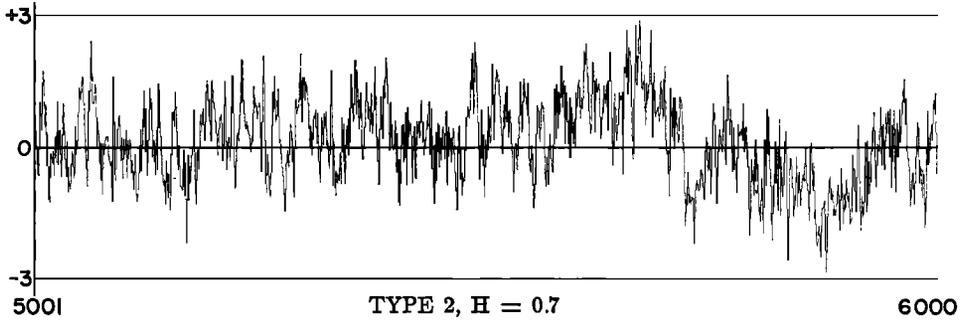
For approximations to fractional noise, $\mathcal{E}[s^{-1} \Delta X^*]^2$ is $\sim C_H s^{2H-2}$ when s is large, but $\mathcal{E}X^2$ need not be $\sim C_H$. Therefore, for the sake of generality, we shall write

$$\mathcal{E}[S^2(t, s)] \perp \sim \mathcal{E}X^2 - C_H s^{2H-2}$$

The expectation $\mathcal{E}[S^2(t, s)]$ increases with s , its limit value for $s \rightarrow \infty$ being $\mathcal{E}X^2$. The increase



Frieze 12. The fourth one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.



Frieze 6. The sixth one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

is faster when H is small than when H is large. For example, when H is nearly 1, $\mathcal{E}[S^2(t, s)]$ remains much smaller than $\mathcal{E}X^2$, even when s is already large. When H is nearly 0, on the contrary, $\mathcal{E}[S^2(t, s)]$ rapidly becomes indistinguishable from $\mathcal{E}X^2$.

When the values of a process cluster tightly around their sample average, one is tempted to say that this process does not fluctuate violently. With such a definition, we see that among fractional noises with common values for $\mathcal{E}X^2$ and C_H , the noises with a high value of H can be considered the least violently fluctuating. This conclusion is contrary to the conclusion of the preceding section. In a later section, on Student's distribution, the two conflicting definitions of violent fluctuation will be reconciled.

CORRELATION BETWEEN PAST AND FUTURE
SAMPLE AVERAGES

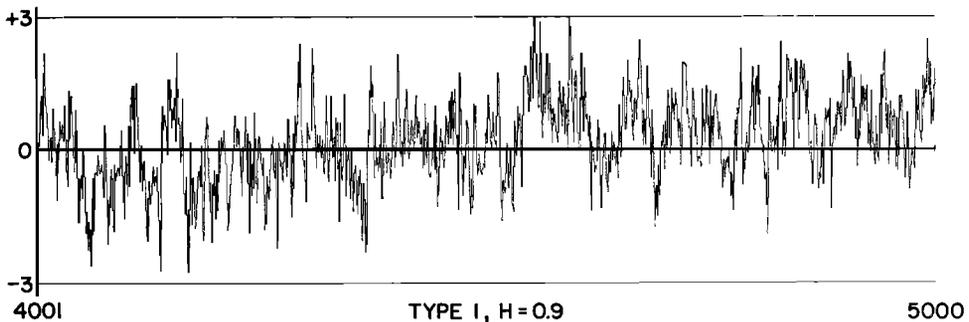
The question that will be solved in this and the next sections has great historical and conceptual interest. It brings us back to the basic

observation of the protohydrologist, Joseph. Even when two successive averaging intervals are both long (seven years!) Joseph observed that the corresponding averages of the flow of the River Nile can differ greatly from each other and therefore from their common expectation [Mandelbrot and Wallis, 1968].

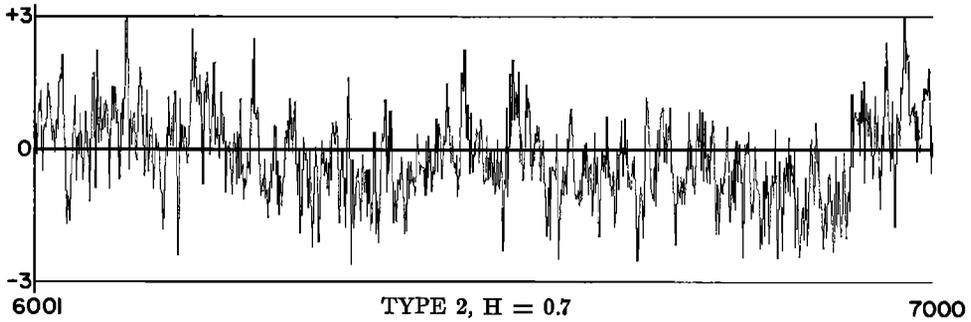
To simplify notation with no loss of generality, we shall select the present moment as origin of time. The two averages in question will be the 'past average' over P years, namely $(1/P) \sum_{u=1-P}^0 X(u)$ and the 'future average' over F years, namely $(1/F) \sum_{u=1}^F X(u)$. In this section, following an argument in Mandelbrot and Van Ness [1968], we shall evaluate the correlation

$$\frac{\mathcal{E} \left\{ \left[\frac{1}{P} \sum_{u=1-P}^0 X(u) \right] \left[\frac{1}{F} \sum_{u=1}^F X(u) \right] \right\}}{\mathcal{E} \left[\frac{1}{P} \sum_{u=1-P}^0 X(u) \right]^2 \mathcal{E} \left[\frac{1}{F} \sum_{u=1}^F X(u) \right]^2}^{1/2}$$

The numerator can be rewritten as



Frieze 13. The fifth one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.



Frieze 7. The seventh one-ninth of a sample of 9000 values of a Type 2 approximation to the fractional noise with $H = 0.7$.

$$\frac{1}{2PF} \left\{ \varepsilon \left[\sum_{u=1}^F X(u) \right]^2 - \varepsilon \left[\sum_{u=1}^0 X(u) \right]^2 - \varepsilon \left[\sum_{u=1}^F X(u) \right]^2 \right\} = \frac{C_H}{2PF} \{ (P + F)^{2H} - P^{2H} - F^{2H} \}$$

The denominator, in turn, becomes

$$\left\{ \frac{C_H}{P^2} P^{2H} \cdot \frac{C_H}{F^2} F^{2H} \right\}^{1/2} = \frac{C_H}{FP} (PF)^H$$

Hence, the correlation equals

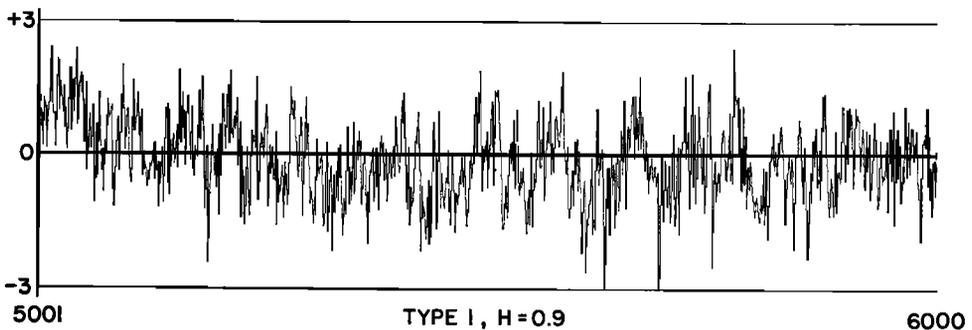
$$\frac{(P + F)^{2H} - P^{2H} - F^{2H}}{2(PF)^H} = \frac{(1 + F/P)^{2H} - 1 - (F/P)^{2H}}{2(F/P)^H}$$

In the classical white noise case, $H = 0.5$, this correlation vanishes, as was expected, since in that case past and present are statistically independent. When $H \neq 0.5$, on the contrary,

the correlation is nonvanishing, save in the special asymptotic cases where $P \rightarrow \infty$ whereas $F/P \rightarrow 0$, or conversely. When, for example, $P \rightarrow \infty$ and $F \rightarrow \infty$, whereas the ratio P/F remains fixed, the correlation between past and future remains unchanged. This brings us to an issue discussed in detail in *Mandelbrot and Wallis* [1968]. We mentioned there that the 'Brownian domain of attraction' for random processes is characterized by *three* properties, namely the law of large numbers, the central limit theorem, and finally the asymptotic independence between past and future. In the case of fractional noises with $H \neq 0.5$, the third property is clearly not satisfied, thus confirming that fractional noises do *not* belong to the Brownian domain of attraction.

Looking at the case $H > 0.5$ more closely, we see that the correlation between past and future is *positive*, and that it increases from 0 to 1 as H increases from 0.5 to 1. This confirms that *fractional noises with $H > 0.5$ are persistent*, such persistence increasing with H .

When $H < 0.5$, on the contrary, the correla-



Frieze 14. The sixth one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.

tion between past and future averages is *negative* and decreases from 0 to -0.5 as H decreases from 0.5 to 0 . This confirms that, in fractional noises with $H < 0.5$, large positive values tend to be followed by large negative values, and vice versa.

In the preceding calculations of correlations, we assumed that M is large, say larger than $P + F$. When M is finite, and $P + F$ much exceeds M , complicated corrections become needed. Asymptotically, if $(P + F)/M \rightarrow \infty$, the behavior characteristic of the Brownian domain of attraction must take over, so that past and future become statistically independent.

DISTRIBUTION OF THE DIFFERENCE BETWEEN PAST AND FUTURE SAMPLE AVERAGES

We shall now study the distribution of the quantity $\Delta_{F,P}$ defined by

$$\Delta_{F,P} = (1/F) \sum_{u=1}^F X(u) - (1/P) \sum_{u=1-P}^0 X(u)$$

We shall begin with mathematics and continue with computer simulations.

The vector of coordinates $\Sigma_{u=1}^F X(u)$ and $\Sigma_{u=1-P}^0 X(u)$ is a bivariate Gaussian vector of zero expectation. Therefore, $\Delta_{F,P}$ is a Gaussian random variable of zero expectation. Its variance is equal to

$$\begin{aligned} \mathcal{E}(\Delta_{F,P}^2) &= \frac{1}{F^2} \mathcal{E} \left[\sum_{t=1}^F X(t) \right]^2 + \frac{1}{P^2} \mathcal{E} \left[\sum_{t=1-P}^0 X(t) \right]^2 \\ &\quad - \frac{2}{FP} \mathcal{E} \left[\sum_{t=1}^F X(t) \sum_{t=1-P}^0 X(t) \right] \\ &= \frac{1}{F^2} \mathcal{E} \left[\sum_{t=1}^F X(t) \right]^2 + \frac{1}{P^2} \mathcal{E} \left[\sum_{t=1-P}^0 X(t) \right]^2 \end{aligned}$$

$$\begin{aligned} &= \frac{1}{FP} \left\{ \mathcal{E} \left[\sum_{t=1-P}^F X(t) \right]^2 - \mathcal{E} \left[\sum_{t=1}^F X(t) \right]^2 \right\} \\ &= C_H \{ F^{2H-2} + P^{2H-2} \\ &\quad - (PF)^{-1} [(P + F)^{2H} - P^{2H} - F^{2H}] \} \end{aligned}$$

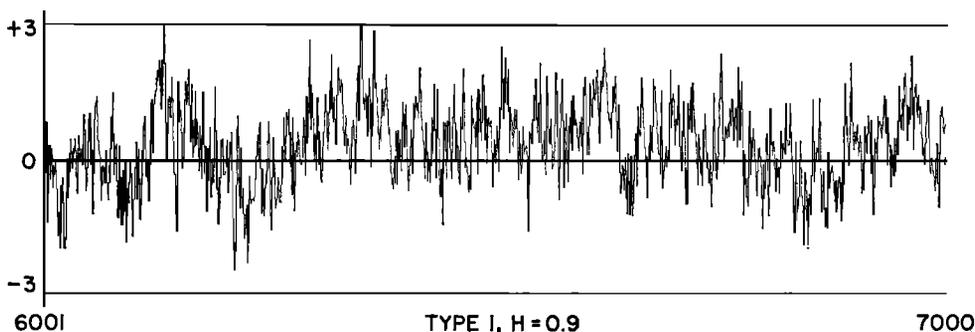
As expected, this is a monotonic decreasing function of both P and F . As $P \rightarrow \infty$ and $F \rightarrow \infty$, past and future averages both tend to zero, and so does the difference $\Delta_{F,P}$ between them. If $P \rightarrow \infty$ while F is finite, the past average tends to zero and

$$\mathcal{E} \Delta_{F,P}^2 \rightarrow \mathcal{E} \left[F^{-1} \sum_{t=1}^F X(t) \right]^2 = C_H F^{2H-2}$$

similarly, when $F \rightarrow \infty$ while P remains finite. When F and P are both finite, $\mathcal{E} \Delta_{F,P}^2$ is a function of P and F , with C_H and H as parameters.

Let us examine the behavior of $\mathcal{E}(\Delta_{F,P}^2)$ when H takes one of the simple values, 0.5 , 0 , and 1 . In the classical white noise case $H = 0.5$, $\mathcal{E}(\Delta_{F,P}^2)$ simply becomes $C_H(F^{-1} + P^{-1})$. This was to be expected, since $\Delta_{F,P}$ is then the difference between two independent variables, of variances P^{-1} and F^{-1} , respectively. Near $H = 0$, $\mathcal{E}(\Delta_{F,P}^2)$ behaves like $C_H[F^{-2} + P^{-2} + (FP)^{-1}]$.

For H very near 1 ($H = 1$ itself being excluded), the constant C_H must be chosen very large, but one can write $C_H = [2(1 - H)]^{-1} C'$ with C' finite. Some algebraic manipulation then yields



Frieze 15. The seventh one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.

$$\begin{aligned} \varepsilon(\Delta_{F,P}^2) = C' & \left[-\left(1 + \frac{F}{P}\right) \log \frac{F}{P} \right. \\ & \left. + \frac{(1 + F/P)^2}{F/P} \log \left(1 + \frac{F}{P}\right) \right] \end{aligned}$$

For $F = P$, this reduces to $4C' \log 2$.

The remark at the end of the preceding section is also valid here: When dealing with approximations to fractional noise, the above formulas for $\varepsilon(\Delta_{F,P}^2)$ require modification.

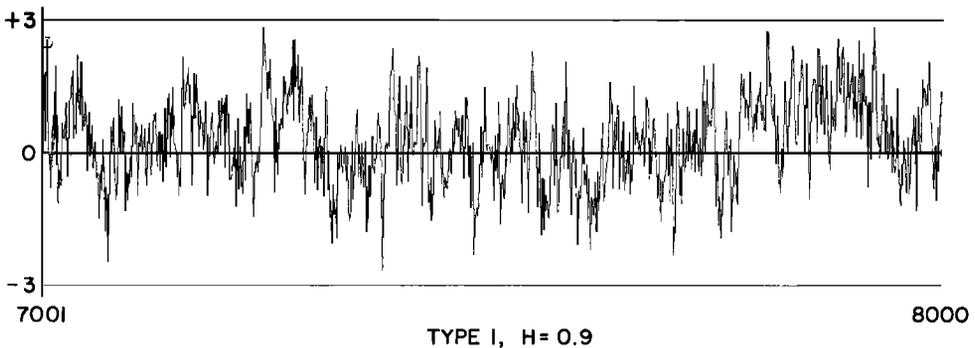
Several simulation experiments were carried out to illustrate the distribution of $\Delta_{F,P}$. The 'horizon' F was kept fixed at $F = 50$, whereas the available past P was varied from 5 to 100. We worked with samples of various Type 1 fractional noises, each of which contained 9000 sample values and had been normalized to have zero mean and unit variance. The number of subsamples of length $P + F$, extracted from each sample, was either 60 or 120, as marked.

Figure 3 (p. 239) illustrates $\Delta_{F,P}$ for Type 1 approximate fractional noises, with various values of the parameter H and a common parameter $M = 10,000$. The curve marked 'white noise' corresponds to the case where $\Delta_{F,P}$ is the difference between two independent Gaussian random variables. The other curves correspond to past and future averages that are correlated, either positively, as when $H > 0.5$, or negatively, as when $H < 0.5$. Keeping P fixed, we see that the average of $|\Delta_{F,P}|$ rapidly increases when H is increased. Keeping H fixed, we see that the average of $|\Delta_{F,P}|$ decreases when P is increased, such decrease being less rapid with $H = 0.9$ than with either $H = 0.7$ or $H = 0.6$. With $H = 0.9$, an increase of P beyond $P =$

50 hardly decreases the uncertainty about the future sample average.

Another important issue was tackled in the experiment leading to Figure 4. The top and bottom curves were reproduced from Figure 3 to emphasize that the middle curve, which corresponds to a Type 1 fractional noise with $H = 0.7$ and $M = 20$, starts close to the upper curve but then drifts rapidly towards the lower curve. This means that the fractional noise with $H = 0.7$ and $M = 20$ has short-run properties nearly identical to those of the fractional noise with $H = 0.7$, $M = 10,000$, whereas its long-run properties are nearly identical to those of white noise.

We wish at this point to digress from dry mathematics and experimental results to explain the specific importance of this last result in hydrology. Statistics, like the test of science (pure and applied), is ruled by the so-called 'Occam's razor,' which states that '*entia non sunt multiplicanda praeter necessitatem*,' and freely translates as 'use the simplest model you can get away with' (because it is not disproved by the data). Let us, in this light, pick a sample of 50 values, known to have been generated by a Type 1 fractional noise with $H = 0.7$ and a very large value of M , and let us try to fit this sample with a fractional noise with $H = 0.7$ and the smallest possible value of M . The quality of such fit may be measured by the $R(t, s)/S(t, s)$ statistic, to be discussed in Part 2 of this paper. When samples shorter than 50 are considered, and when $H = 0.7$, a fractional noise having a small value of M provides a bad approximation to the postulated process, which has a large M . The quality of approximation first



Frieze 16. The eighth one-ninth of a sample of 9000 values of a Type 1 fractional noise with $H = 0.9$.

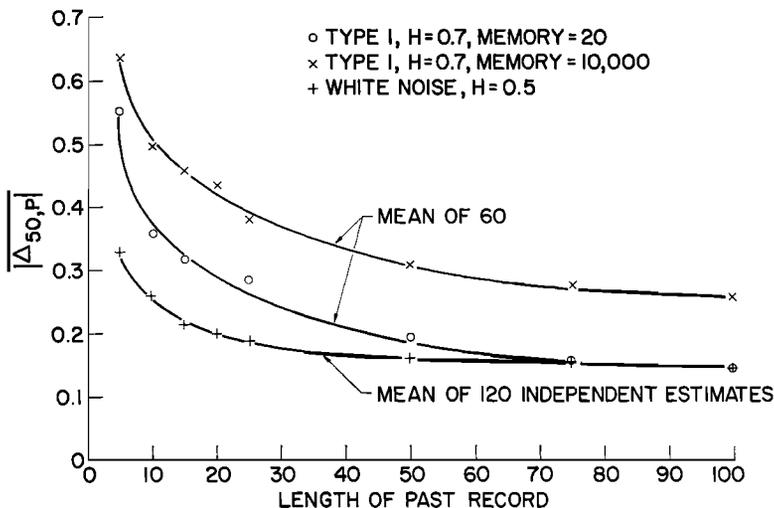


Fig. 3. Given various Type 1 approximate fractional noises (as marked), this figure shows the behavior of $\Delta_{50,P}$, defined as the difference between the sample average over the next $F = 50$ years, and the sample average of the past P years. The abscissa is P . Our program yielded a sample average of $|\Delta_{50,P}|$, which was plotted as ordinate. The square root of the sample average $\Delta_{50,P}^2$ would have been easier to compare with the mathematical formulas in the text. However, the two quantities are approximately proportional, so that our program was not modified. The reader is, however, warned against possible misinterpretation of this and the following figure.

improves as M increases but ceases to improve if M goes beyond 20. Thus, with short records, a memory of 20 is suggested by a naive application of Occam's razor. Another statistic, the sample covariance of a record of 50 data, is also likely to be adequately fitted by the covariance of a moving average process of memory 20.

It would be unreasonable, however, to try to use this best fitted process to forecast the sample average over the next 50 or 100 years. The behavior of the middle curve in Figure 4 shows that a 20-year moving average is *dreadful* for the purpose of forecasting $(1/F) \sum_{u=1}^F X(u)$. It is hardly better than a process taking independent values.

This kind of deficiency is very serious whenever one deals with an underlying process known to have very strong long-run effects. In such cases, to be able to extrapolate reasonably from the past P to the future F years, one must take account of effects that are fully developed only when samples of length $P + F$ are available.

To account for such effects, information available in records of the past is unavoidably deficient, so that one must add some extra information borrowed from other data believed to resemble the data being studied. In the present

context, this general precept can be implemented as follows. If there is evidence that one stream is characterized by $H = 0.7$, it is better to evaluate $\Delta_{F,P}$ 'as if' the past had been generated by a process with $H = 0.7$ and a very large M , rather than generated by the simplest model with which the available past sample can reasonably be fitted.

In conclusion, the rare cases when very long-term storage is envisioned are not the only ones where there is need for considering long-run effects. Such need is already present in the innumerable designs that imply a forecast of the next 50 years on the basis of a knowledge of the last 50 years.

THE COUNTERPART OF STUDENT'S DISTRIBUTION

The preceding discussion of $\Delta_{F,P}$ implied the unrealistic assumption that one knows the value of the constant C_H . In fact, in practical situations, one knows neither $\mathcal{E}X$, nor $\mathcal{E}X^2$, nor any other theoretical expectation. One only knows the quantities that can be computed on the basis of a sample of P past values of $X(t)$. We shall in this section study $\Delta_{F,P}$ as a proportion of the standard deviation S_P of the last P values of the process. The study of $\Delta_{F,P}/S_P$ is

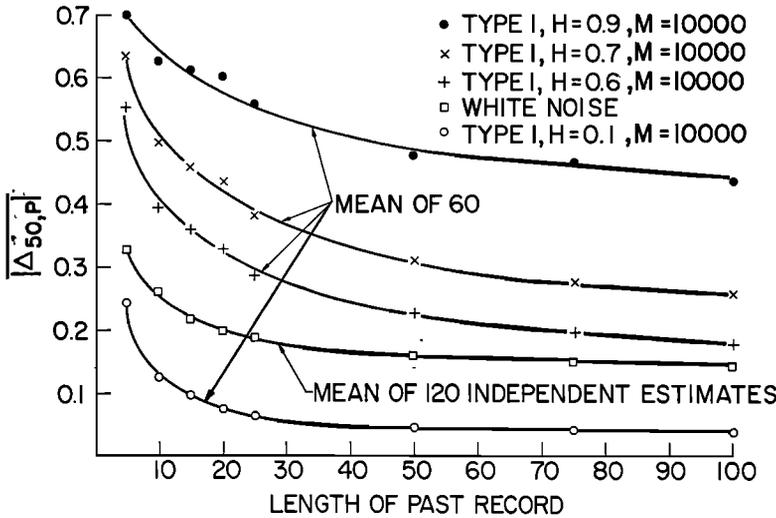


Fig. 4. The effect of several combinations of the memory M and of the exponent H upon the behavior of the sample average of $|\Delta_{50,P}|$. (See caption of the preceding figure for a warning concerning this choice of an average.) Considering a fractional noise with $H = 0.7$, and $M = 20$ from the viewpoint of $|\Delta_{50,P}|$, it is apparent that its short-run properties resemble those of a fractional noise with $H = 0.7$ and $M = 10,000$, whereas its long-run properties resemble those of white noise.

comparatively complicated. We present it here because it will be needed in applications and because it is useful in dispelling the paradox of the least and most erratically fluctuating among fractional noises. However, the discussion that follows will not be used in the presently written papers of this series, and the reader may directly proceed to Part 2.

In the case of independent Gauss random processes, the distribution of $\Delta_{F,P}/S_P$ is related to a distribution due to Student (who had to hide his real name, William Gossett, to prevent the competitors of Guinness Breweries from finding out that statistics could help brewers make money). 'Student' formed the ratio

$$t_p = \left[\varepsilon X - P^{-1} \sum_{u=1}^0 X(u) \right] [P - 1]^{0.5} / S_P$$

The distribution of the random variable t_p , called 'Student's distribution with $P-1$ degrees of freedom,' is described today in every book of statistics. Unless P is large, this distribution is extremely long tailed, that is, t_p can take very large values with appreciable probability. This feature confirms the intuitive feeling that large relative errors are easily made unless P is quite large.

A corollary of Student's result is that

$$\frac{\Delta_{F,P}}{S_P} = \frac{\left[P^{-1} \sum_{u=1}^0 X(u) - \varepsilon X \right]}{S_P} \quad \frac{\left[F^{-1} \sum_{u=1}^F X(u) - \varepsilon X \right]}{S_P}$$

can be written as the difference of two independent random variables, namely as $t_p(P - 1)^{-0.5} - t_p(F - 1)^{-0.5}$.

Now replace the independent Gaussian process ($H = 0.5$) by a fractional noise with either $0 < H < 0.5$ or $0.5 < H < 1$. The two ratios

$$S_P^{-1} \left\{ P^{-1} \sum_{u=1}^0 X(u) - \varepsilon X \right\}$$

and

$$S_P^{-1} \left\{ F^{-1} \sum_{u=1}^F X(u) - \varepsilon X \right\}$$

are no longer independent, and the distribution of their difference $\Delta_{F,P}/S_P$ becomes hard to derive. To simplify the continuation of this discussion, we shall make the additional assumption that $F = P$. A heuristic argument suggests for large P : (a) that $\Delta_{F,P}$ and S_P are nearly independent, because $\Delta_{F,P}$ depends primarily upon low frequency effects, whereas S_P depends primarily upon high frequency effects, and (b)

TABLE 1.

<i>H</i>	<i>P</i>	<i>V</i>	<i>H</i>	<i>P</i>	<i>V</i>	<i>H</i>	<i>P</i>	<i>V</i>
0.1	5	0.25	0.5	5	0.50	0.7	5	0.65
	10	0.15		10	0.33		10	0.48
	30	0.05		30	0.18		30	0.32
	50	0.03		50	0.14		50	0.27
	100	0.02		100	0.10		100	0.21
	300	0.01		300	0.06		300	0.15
0.3	5	0.38	0.6	5	0.57	0.9	5	0.82
	10	0.23		10	0.40		10	0.66
	30	0.10		30	0.24		30	0.51
	50	0.07		50	0.20		50	0.47
	100	0.04		100	0.15		100	0.41
	300	0.02		300	0.09		300	0.35

that the variance of $\Delta_{F,P}/S_P$ is roughly equal to the ratio of the variances of its numerator and denominator, namely

$$\begin{aligned} \varepsilon \left[\frac{\Delta_{F,P}}{S_P} \right]^2 &\sim \frac{\varepsilon(\Delta_{F,P}^2)}{\varepsilon(S_P^2)} \\ &\sim \frac{C_H(4P^{2H-2} - 4^H P^{2H-2})}{C_H(1 - P^{2H-2})} \perp = \frac{4(1 - 4^{H-1})}{P^{2-2H} - 1} \end{aligned}$$

We shall designate the last term to the right by $2V^2(P, H)$. When $H = 0.5$, $V^2(P, 0.5) = 2(P - 1)^{-1}$, which is indeed the exact value for $F = P$ of the variance of $t_P(P - 1)^{-0.5} - t_P(F - 1)^{-0.5}$. For other values of H and P , numerical values of V are tabulated in Table 1.

A third prediction of the heuristic argument is (c) that if H is changed from $H = 0.5$ to some other value, $\Delta_{F,P}/S_P$ is unchanged in the form of its distribution but is simply multiplied by $V(P, H)/V(P, 0.5)$. Using computer simulation, it was established that this last prediction constitutes a good first approximation.

If one disregards the behavior of a tail in which 1% of the total probability is concentrated, the rescaled variable $[\Delta_{F,P}/S_P] [V(P, 0.5)/V(P, H)]$ is roughly distributed as the difference $t_{P'}(P - 1)^{-0.5} - t_{P''}(P - 1)^{-0.5}$, where $t_{P'}$ and $t_{P''}$ are two independent Student's random variables.

This is a good point to return to the discussion of which are the most 'violently fluctuating' among fractional noises. We saw earlier that, for given values of P, C_H , and εX^2 , $\varepsilon[\Delta_{F,P}^2]$ increases with H , showing, as expected, that fractional noises with a high value of H are the most violently fluctuating of all. We also saw, however, that S_P decreases as H increases. Were the 'violence' of fluctuations to be measured by quantities that increase with S_P , such a de-

pendence of S_P upon H would have been paradoxical. Actually, however, the 'violence' of fluctuations is best measured by quantities that decrease as S_P increases. An example is $\Delta_{F,P}/S_P$, in which S_P enters as the denominator. No paradox is therefore present. Not only does the ratio $\Delta_{F,P}/S_P$ increase with H , but such increase is due to the combined effects of an increase of the numerator and a decrease of the denominator.

The distribution of $t_{P'}(P - 1)^{-0.5} - t_{P''}(P - 1)^{-0.5}$ is not given in standard statistical tables. Fortunately, it is not needed when P is very large. Indeed, as $P \rightarrow \infty$, it is known that t_P tends towards a Gaussian of zero mean and unit variance, so that the limit of $t_P(P - 1)^{-0.5} - t_P(F - 1)^{-0.5}$ is itself approximately Gaussian with zero mean and variance $(P - 1)^{-1} + (F - 1)^{-1}$. One will then want to know how large P must be, in order that t_P be 'nearly Gaussian.' For a preliminary investigation, it was sufficient to require that t_P approximate its Gaussian limit with a relative error that is 'small' (say, 5% or 10%) except in a small number (say, 1%) of cases. By inspecting the table of Student's distribution, one finds that, except for 1% of cases, t_P is within 10% of its Gaussian limit whenever $P > 20$, and that t_P is within 5% of its Gaussian limit whenever $P > 40$. Thus, tables of the Gaussian may suffice in the first approximation when P equals one 'lifetime' of about 50 years.

The practical application of the above results would then proceed as follows. Given a large P and a small threshold p , tables of the Gaussian distribution give the value that $\Delta_{F,P}/S_P$ will exceed with the probability p in the case $H = 0.5$. If $H \neq 0.5$, the value obtained above is simply to be multiplied by V .

(For the reader insisting on mathematical precision, the above statement about nearly Gaussian distributions may be reexpressed as follows: Let $g(p)$ be the threshold exceeded with the probability p by a Gaussian random variable of zero mean and unit variance, and let $x_P(0)$ be the corresponding threshold for Student's t_P . To ensure that $x_P(p)/g(p)$ lies between 1 and 1.1 for all p between 0.005 and 0.5, it suffices that $P > 20$, and by symmetry. To ensure that $x_P(p)/g(p)$ lies between 1 and 1.05 for all p between 0.005 and 0.5, it suffices that $P > 40$ and by symmetry.)

REFERENCES. See Page 267.

(Manuscript received September 16, 1968.)