Water Resources Research 5, 1969, 228-241.

Computer experiments with fractional Gaussian noises. Part 1: Sample graphs, averages and variances (M & Wallis 1969a)

• ***Chapter foreword.*** The words "sample graphs" were added to this chapter's title to emphasize the extent (unusual in 1969) to which the original paper relied on the eye to help train new intuition. The story is told in Section 4.3.2 of Chapter H8.

A matter of layout. To accommodate better the many illustrations in this paper, many are *not* printed *after* the first reference to them, but *before*.•

✦ **Abstract.** Fractional Gaussian noises are a family of random processes with long statistical dependence. This means that their values at instants of time very distant from each other have a correlation that is small but nonneglibible. Mathematical analysis has shown that the interdependence in these processes has precisely the intensity required to satisfy Hurst's law $R/S \sim \delta^H$; therefore, these processes provide a good mathematical model of long-run effects in hydrological and geophysical records. This paper goes beyond mathematical analysis by introducing computer simulations. With their help, the results of the analysis are illustrated, extended and restated in a form suitable for practical use. This chapter presents and discusses the shape of the sample functions and the correlation between the past and the future averages. Chapter H13 tackles other aspects of the same process. Mathematical details are withheld until Chapter H14.{P.S. 2000: they are treated in greater detail in Chapter 11.} ✦

USING A COMPUTER AND A GRAPHIC PLOTTER AS TOOLS, we have carried out simulation experiments concerning the sample behavior of "fractional Brownian motions" and "fractional Gaussian noise;" the results are reported in this paper and the next two. M 1965h{H9} and M

& Wallis 1968{H10}, proposed fractional Gaussian noises as possible models of hydrological records. Full empirical evidence for their quality as models will be provided in M & Wallis 1969b{H27}.

INTRODUCTION

As defined in M & Van Ness 1968{H11}, fractional Brownian motions and fractional Gaussian noises (FBM and FGN, respectively) generalize "Brownian motion" and "white Gaussian noise," respectively. A white Gaussian noise sample is shown in Figure 1. It is a "sequence of independent Gaussian random variables."

The term "fractional noise" is justified by considerations from spectral theory. (The bulk of this paper can be followed even if this justification is not fully appreciated.) Classically, a white noise is defined as a random process having a spectral density independent of the frequency f; it is convenient to write it as f^{-0}. As a result, the integral and derivative of white noise, and its repeated integrals or derivatives, all have spectral densities of the form f^{-B}, where $B/2$ is as an integer. On the contrary, fractional noises is defined as having a spectral density of the same form f^{-B}, where $B/2$ is *not an integer*. This explains the term "fractional white noise." By repeated integration or differentiation, one can restrict k to any interval of unit length. For reasons that will become apparent later, B will be written as $1-2H$, where H, the main parameter of a fractional noise, will vary between 0 and 1. Classical white noise occurs when $H=0.5$. Despite their role in the above definition, spectral techniques will be the least important among the many techniques used in this paper. Figure 1. •

We employed two approximations to fractional Gaussian noise in our simulations. Our "Type 2" approximation is far less accurate and less important but far easier to define. It is given by the two-parameter moving average Figure 2. •

$$F_2(t \mid H, M) = (H-0.5) \sum_{u=t-M}^{t-1} (t-u)^{H-1.5} G(u) + Q_H G(t).$$

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

$$Q_H = \begin{cases} 0 & \text{if } 0.5 < H < 1, \\ (0.5 - H)\sum_{u=1}^{\infty} u^{H-1.5} & \text{if } 0 < H < 0.5. \end{cases}$$

The parameter M is called the "memory of the process." In theory (M 1965h), one should set $M = \infty$, but in practice this is impossible. 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" approximations are more cumbersome and are postponed until the chapter after next.

For actual records (M & Wallis 1969b), the Gaussian assumption is only an approximation, and the non-Gaussian fractional noises investigated in M & Van Ness 1968 are very important. However, the epithet "Gaussian" can be omitted in this paper; we shall simply speak of "fractional noise."

FORM OF THE SAMPLE FUNCTIONS OF FRACTIONAL GAUSSIAN NOISE

Figure 1 is a short sample of a process of independent Gaussian random variables. A short sample suffices because this process is monotonous and "featureless." It is analogous to the "hum" in electric amplifiers and is often called a "discrete-time white noise." It can also be considered a discrete-time fractional noise with $H = 0.5$.

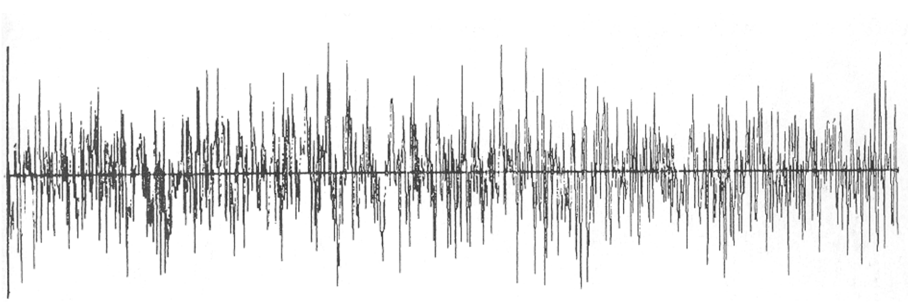


FIGURE C12-1. A sample of 1000 values of a discrete white noise of zero mean and unit variance. Alternative names are sequence of independent Gaussian variables and discrete fractional Gaussian noise with $H = 1/2$.

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

Many of the remaining illustrations of this paper concern two very long samples of FGN with $1/2 < H < 1$. It would be self-defeating to print them in a tight format. Furthermore, this paper hopes to create in the reader's mind a strong intuition of the form of FGN. We have, therefore, formatted the text so that two values of H can be compared on many of the pages. This innovation creates what we call "friezes," each containing 1000 values of a total of 9000.

Friezes 1 to 8 run along the page tops and carry successive samples of a moderately nonwhite fractional noise with $H=0.7$. Friezes 9 to 16 run along the page bottoms and carry a strongly nonwhite fractional noise with $H=0.9$. Whenever $H>0.5$, a fractional noise is richer than 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 effects of such low frequency terms.

We wish to encourage comparison of our artificial series with the natural records with which the reader is concerned. To be meaningful, the comparison must involve the same degree of local smoothing of high frequency "jitter" in both cases.

The reader must not be ashamed to listen to "gut feelings." When the artificial series "feel" very different from the natural records, then fractional noises are probably not an appropriate model. If the artificial series "feel" close to nature but not quite right, then perhaps look at a different value of H. When the artificial series "feel" right, then the reader should

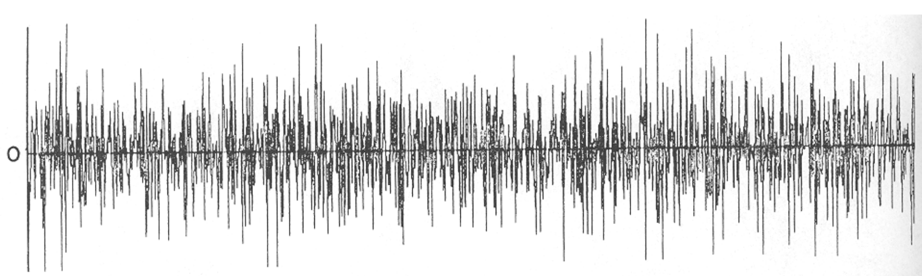


FIGURE C12-2. A sample of 1000 values of a Type 1 FGN with $H=0.1$.

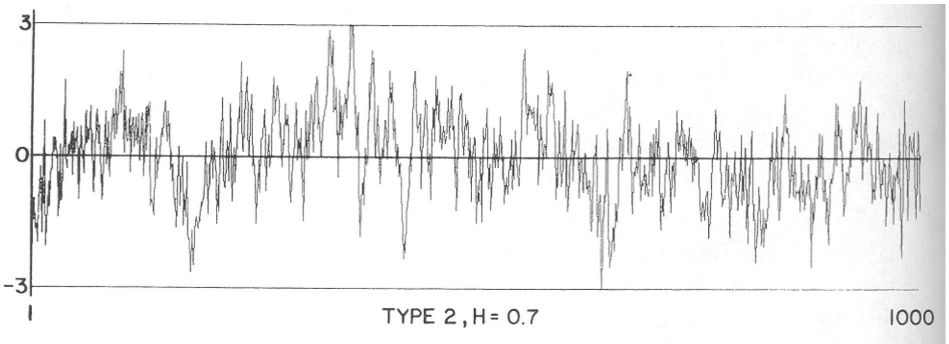
proceed to formal statistical tests of fit. Such formal tests are indispensable, but a statistical test by necessity focuses upon a specific aspect of a process, whereas the eye can often balance and integrate various aspects. We emphasize that formal testing and visual inspection should be combined, but it is best to start with the inspection.

To cite our own experience, visual inspection helped us to establish that Hurst was mistaken in his belief that the H coefficient of hydrological records is "typically" near 0.7. Indeed, the sample noise with $H=0.7$, which we had plotted first, seemed much too smooth and too 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 $H=0.9$ with other values, such as $H=0.5$ itself, being possible. This conclusion has since been confirmed by direct testing, as seen in M & Wallis 1969b{H27}.

FRACTIONAL GAUSSIAN NOISE WITH $0.5 < H < 1$ AND APPARENT "CYCLES;" THE "ONE-THIRD" RULE

Examination of the friezes reveals a striking characteristic of fractional Gaussian noise: its sample functions exhibit an astonishing wealth of "features" of every kind, including features that bear an irresistible resemblance to trends and cyclic swings of various frequencies. In some samples, the swings are rough and far from periodic, but in other samples they look periodic. However, the wavelength of the longest apparent cycle markedly depends on the total sample size. If one's attention is deliberately focused on a shorter portion of these friezes, shorter cycles become visible. At the other extreme, from our original plots of these friezes as strips of 3000 time units, it was difficult to avoid the impression 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 real in the sense that something present in human perceptual mechanisms brings most observers to "recognize" the same cyclic behavior. Such cycles are useful to describe the past, but have no predictive value for the future. Analogous remarks were made by Keynes 1939 about economic cycles. This is not surprising, since the resemblance between economic and hydrological time series has been often noted.

The "one-third rule." As a rough rule of thumb valid unless the sample size chart is very short, we have found that the longest cycles have a wavelength equal to one third of the sample size. This rule is not only a



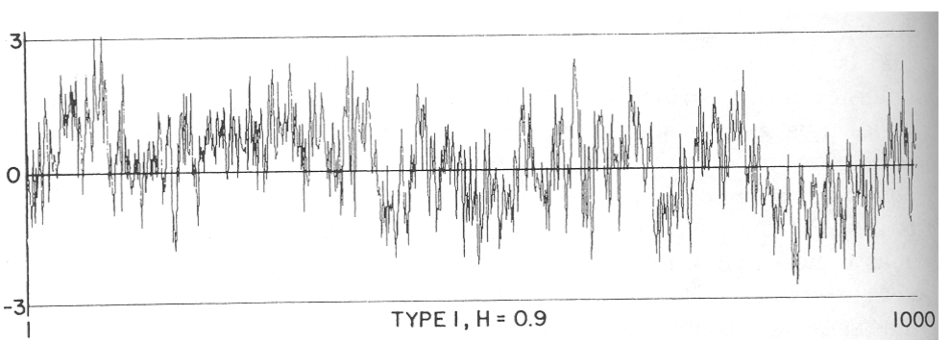
Frieze 1. The first one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$. The overall sample was normalized to have zero mean and unit variance. The Type 2 approximation was selected because a Type 1 approximation with $H = 0.7$ has so much high frequency "jitter" that low frequency effects are illegible.

matter of the psychology of perception, but also a consequence of the self-similarity of fractional noise, as will be seen in the chapter after next.

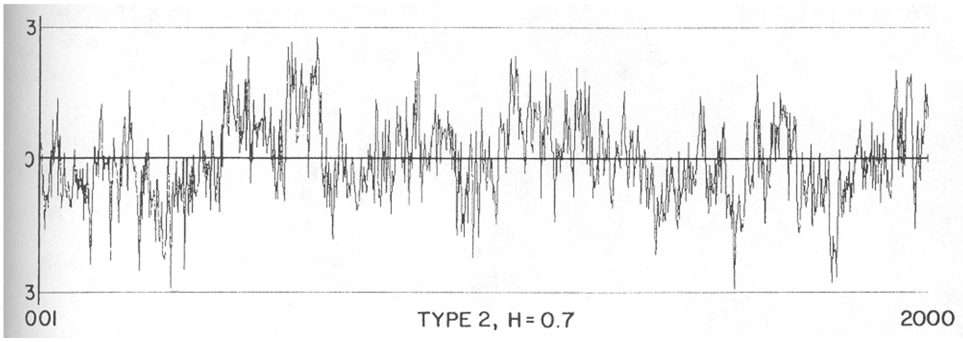
The classical mathematical technique for studying periodicities is spectral or Fourier analysis; this will be examined in Part 2 (next chapter). It suffices to say at this point that spectral analysis *does not* confirm the apparent periodic appearance of fractional noise.

FRACTIONAL GAUSSIAN NOISE, "RUNS" AND HIGH AND LOW "TERMS"

A second and even more striking characteristic of fractional noise with



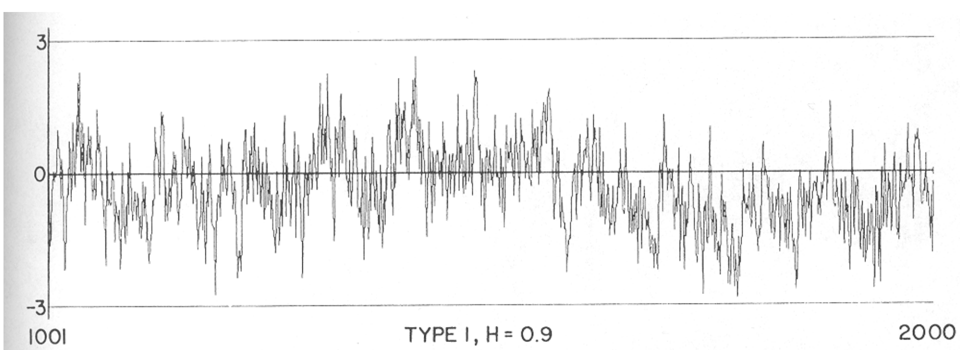
Frieze 9. The first one-ninth of a sample of 9000 Type 1 FGN with $H = 0.9$. The overall rough was normalized to have zero mean and unit variance. When $H = 0.9$, the high frequency "jitter" is too weak to hide the low frequency effects.



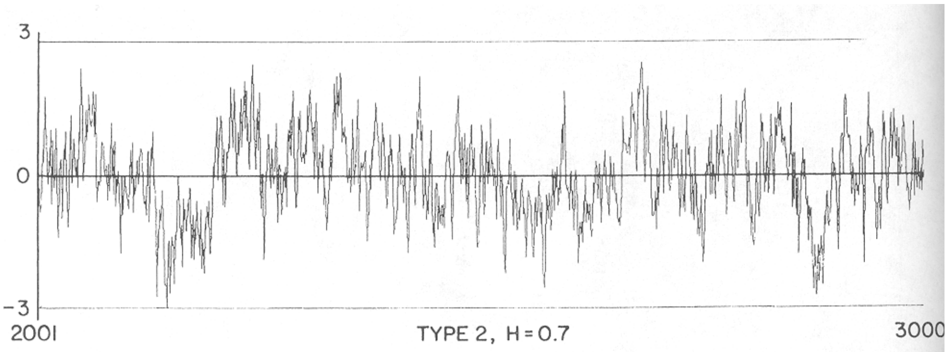
Frieze 2. The second one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$.

$H > 0.5$ 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 years of the Biblical Joseph, the son of Jacob. One is tempted to express this perceptual "persistence" of fractional noise by quantifying the idea of a "run" of dry weather, or a "dry run." For example, a dry run could be defined as a period when precipitation stays below some specified line; a "wet run" is a period when precipitation stays above this line. However, a careful inspection of fractional noise shows many instances where the notion of run describes very poorly the behavior of a time series. 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 three runs? Or is it "really" a single run? Perhaps the short dry run in question could be eliminated by a little smoothing?

Our attempts to define runs reasonably were 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



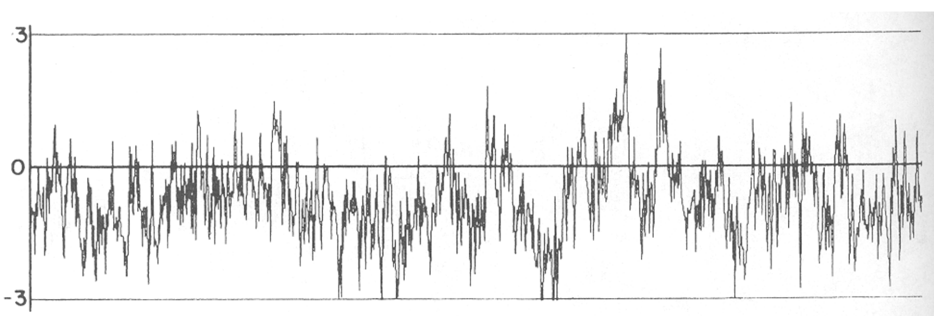
Frieze 10. The second one-ninth of a sample of 9000 Type 1 FGN with $H = 0.9$.



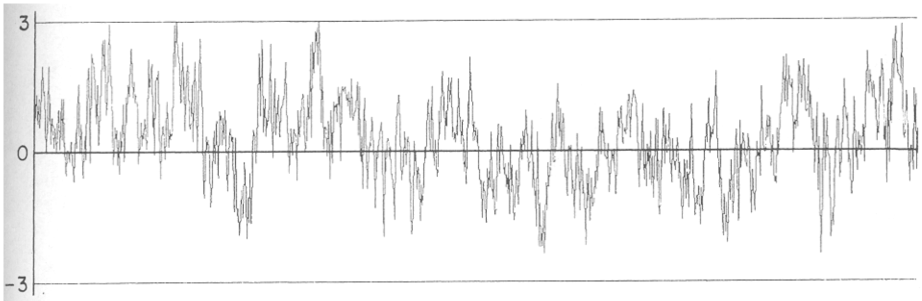
Frieze 3. The third one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$.

various lengths) and scores of redefinitions of "dry" and "wet" (different crossover levels). The distribution 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.

Other authors agree that the notion of a run is too elusive to be useful. One must seek a better way to express 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," of a low term. Mathematically, the average level during a high term may be exceeded by only a few extremes during the preceding or following low term. The sections that follow will suggest several precise ways of expressing the concept of "term."



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



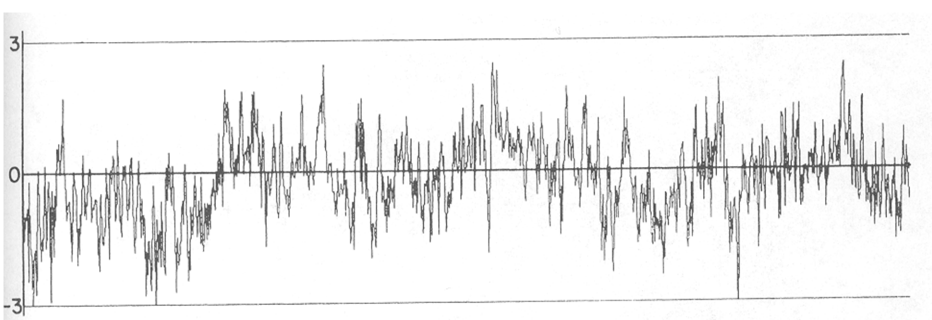
Frieze 4. The fourth one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$.

THE VARIANCE OF SECULAR AVERAGES OF FRACTIONAL NOISE

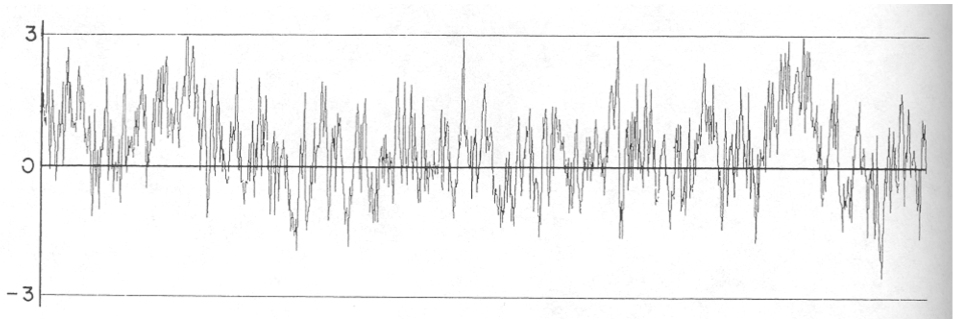
Following the usage of astronomy, an average taken over a time span of the order of 50 or 100 years will be called secular. Successive secular averages of a white noise are practically indistinguishable from one another. On the contrary, let us confirm the perceptual impression that fractional noise with $H > 0.5$ is persistent meaning that averages over successive spans of δ years differ markedly from the population expectation, which is set to 0. To account for this discrepancy in behavior, we shall now evaluate the variance of the secular average about the origin as a function of δ and H.

For simplicity, the argument will not be presented for the Type 1 approximation, but for a different approximation, the discrete-time fractional noise defined by

$$X(t) = B_H(t+1) - B_H(t).$$



Frieze 12. The fourth 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 Type 2 FGN with $H = 0.7$.

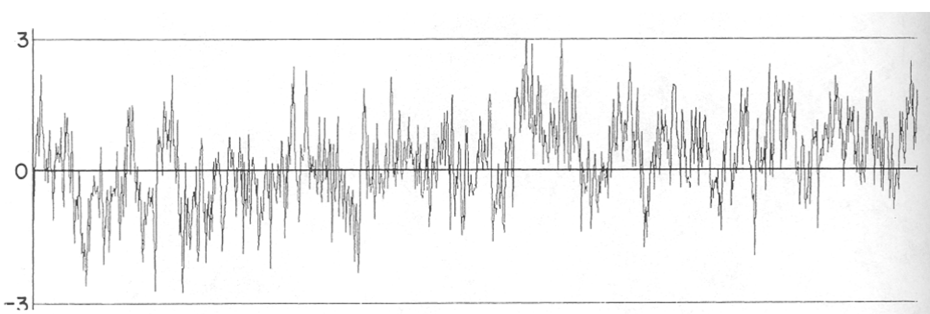
The chapter after next shows that

$$E[\Delta X_\Sigma]^2 = = \left[\sum_{u=1}^{\delta} X(t+u) \right]^2 = E[\Delta B_H]^2 = C_h \delta^{2H},$$

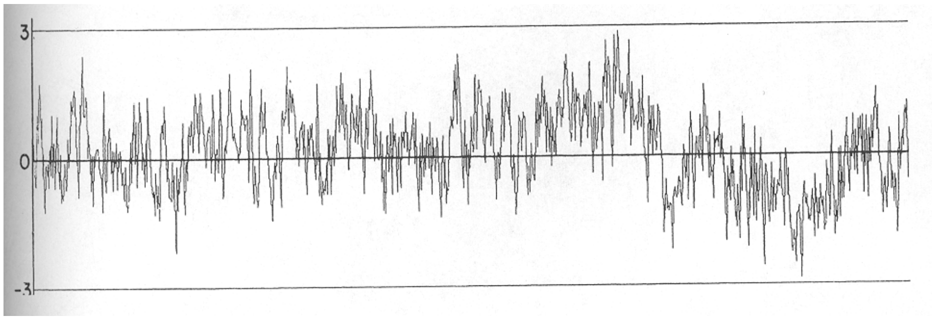
where the symbol E denotes the expectation, and C_H is a positive constant.

In the case of approximate fractional noise, the above formula only applies for "large" δ, where the point at which δ can be called "large" depends upon H and upon the approximation used. This detail is not important to the present discussion. The main fact is that, for large δ, the average of a fractional Gaussian noise has the variance

$$E[\delta^{-1} \Delta X_\Sigma]^2 = E\left[\delta^{-1} \sum_{u=1}^{\delta} X(t+u) \right]^2 = C_H \delta^{2H-2}.$$



Frieze 13. The fifth one-ninth of a sample of 9000 Type 1 FGN 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 $H = 0.7$.

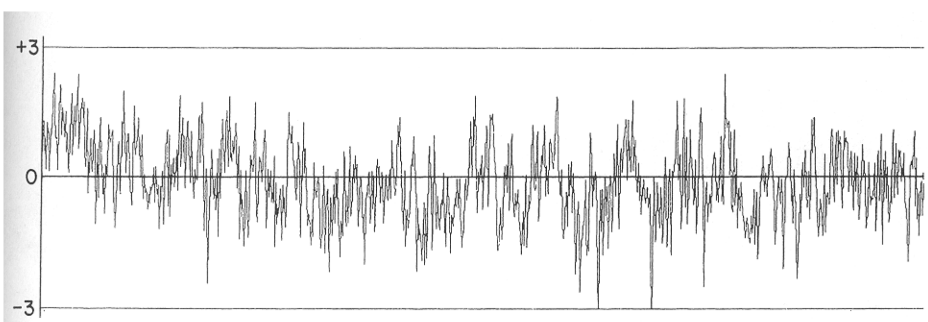
The corresponding standard deviation is $\sqrt{C_H}\,\delta^{H-1}$. This expression always decreases to zero as $\delta \to \infty$, but its rate of decrease depends sharply on H.

To give a numerical example, let $\delta = 100$. Then, we have the following correspondence

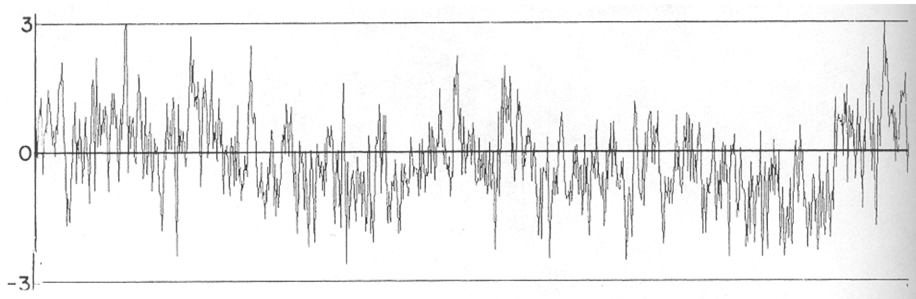
H:	0.1	0.5	0.7	0.9
δ^{H-1}:	0.0016	0.10	0.25	0.62.

That is, δ^{H-1} reduces to a mere $1/10$ in the classical Brownian case $H = 0.5$. It becomes minute when $H = 0.1$. But it is $1/4$ when $H = 0.7$.

When $H = 0.5$, one has a normal rate of decrease of the standard deviation by averaging; otherwise the rate of decrease is "anomalous." When $1/2 < H < 1$, one is tempted to say that this process "fluctuates violently."



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



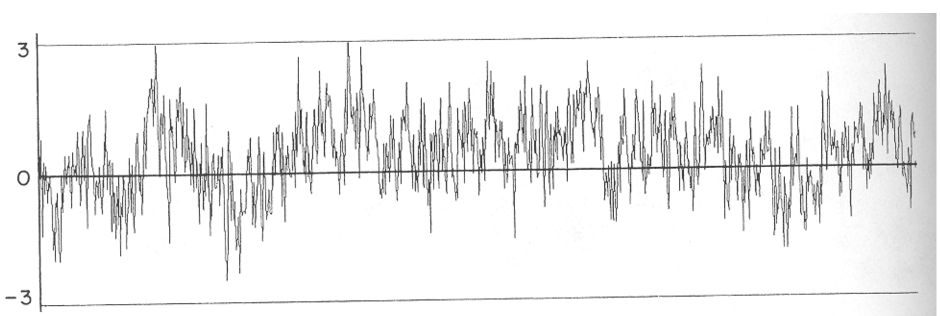
Frieze 7. The seventh one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$.

Given such a definition, we see that *fractional noises with a high value of H are the most violently fluctuating among fractional noises*.

Curves that show how the variance of the secular average depends upon δ, are sometimes called "variance time" curves by statisticians, whose textbooks credit the notion to G. Udny Yule. In fact, these curves are older; for example, Taylor 1921 used them to study turbulent diffusion (see Friedlander & Topper 1961).

THE SEQUENTIAL VARIANCE

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



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

Frieze 8. The eighth one-ninth of a sample of 9000 Type 2 FGN with $H = 0.7$.

$$
\begin{aligned}
S^2(t, \delta) &= \delta^{-1} \sum_{u=1}^{\delta} \left[X(t+u) - \delta^{-1} \sum_{u=1}^{\delta} X(t+u) \right]^2 \\
&= \delta^{-1} \sum_{u=1}^{\delta} X^2(t+u) - \left[\delta^{-1} \sum_{u=1}^{\delta} X(t+u) \right]^2 \\
&= \delta^{-1} \sum_{u=1}^{\delta} X^2(t+u) - \left[\delta^{-1} \Delta X_{\Sigma} \right]^2 .
\end{aligned}
$$

Consequently,

$$
ES^2(t, \delta) = EX^2 - E\{\delta^{-1}[X_{\Sigma}(t+\delta) - X_{\Sigma}(t)]\}^2 .
$$

For discrete-time fractional noise, this becomes

TABLE 1.

H	P	V	H	P	V	H	P	V
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

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

$$ES^2(t, \delta) = C_H - C_H \delta^{2H-2}.$$

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

$$E[S^2(t, \delta)] \sim EX^2 - C_H \delta^{2H-2}.$$

The expectation $E[S^2(t, \delta)]$ increases with δ, its limit value for $\delta \to \infty$ being EX^2. The increase is faster when H is small than it is when H is large. For example, when H is nearly 1, $E[S^2(t, \delta)]$ remains much smaller than EX^2, even when δ is already large. When H is nearly 0, to the contrary, $E[S^2(t, \delta)]$ rapidly becomes indistinguishable from EX^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 EX^2 and C_H, the noises with a high value of H are the least violently fluctuating*. This conclusion is contrary to the conclusion of the preceding section. In a later section, devoted to Student's distribution, the two conflicting definitions of violent fluctuation are reconciled.

CORRELATIONS BETWEEN THE PAST AND THE FUTURE SAMPLE AVERAGES

The question that will be solved in this and the following sections has great historical and conceptual interest. It returns us to the basic observation of the most famous protohydrologist, Joseph.

He observed the corresponding averages of the flow of the River Nile over successive intervals of seven years, and noted that these averages can differ greatly from each other and, therefore, from their common expectation (M & Wallis 1968{H}).

To simplify the notation with no loss of generality, we shall select the present moment as the 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, $(1/F)\sum_{u=1}^{F} X(u)$.

M & Van Ness 1968, show that the correlation between past and future is

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

The numerator can be rewritten as

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

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

The denominator, in turn, becomes

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

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, with $H=0.5$, this correlation vanishes (as was expected) since the past and the present are statistically independent. When $H\neq 0.5$, on the contrary, the correlation is nonvanishing, except in the special asymptotic cases where $P\rightarrow\infty$ while $F/P\rightarrow 0$, or conversely. For example, let $P\rightarrow\infty$ and $F\rightarrow\infty$, while the ratio P/F remains fixed; then the correlation between the past and the future remains unchanged.

This brings us to an issue M & Wallis 1968{H10} discusses in detail. It is mentioned in that paper that the "Brownian domain of attraction" for random process is characterized by *three* properties: the law of large numbers and the central limit theorem, as expected, but also the asymptotic independence between the past and the future. In the case of fractional noises with $H \neq 0.5$, the first two properties are preserved, but 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 the past and the 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*, and that persistence increases with H.

When $H < 0.5$, on the contrary, the correlation between the past and the future is *negative* and decreases from 0 to -0.5 as H decreases from 0.5 to 0. This confirms that, in fractional noise with $H < 0.5$, large positive values tend to be followed by large negative values, and vice versa.

In preceding calculations or correlations, we assumed that M is large, for example, larger than $P+F$. When M is finite, and $P+F$ greatly exceeds M, complicated corrections become necessary. As $(P+F)/M \rightarrow \infty$, the behavior characteristic of the Brownian domain of attraction must become dominant that is, the past and the future become statistically independent.

DISTRIBUTION OF THE DIFFERENCE BETWEEN THE PAST AND THE FUTURE SAMPLE AVERAGES

We shall now study the distributions 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 proceed with computer simulations.

The vector of coordinates $\sum_{u=1}^{F} X(u)$ and $\sum_{u=1-P}^{0} X(u)$ 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

$$E(\Delta_{F,P}^2) = \frac{1}{F^2}E\left[\sum_{t=1}^{F} X(t)\right]^2 + \frac{1}{P^2}E\left[\sum_{t=1-P}^{0} X(t)\right]^2 - \frac{2}{FP}EE\left[\sum_{t=1}^{F} X(t)\sum_{t=1-P}^{0} X(t)\right]$$

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

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

$$= C_H\{F^{2H-2} + P^{2H-2} - (PF)^{-1}[(P+F)^{2H} - P^{2H} - F^{2H}]\}.$$

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

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

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

Let us examine the behavior of $E(\Delta_{F,\,P}^2)$ for a few simple values of H.

In the classical white noise case where $H = 0.5$, $E(\Delta_{F,\,P}^2)$ simply becomes $C_H(F^{-1} + P^{-1})$. This is expected, since in this case, $\Delta_{F,P}$ is the difference between two independent variables whose variances are P^{-1} and F^{-1}, respectively.

Near $H = 0$, $E(\Delta_{F,\,P}^2)$ behaves like $C_H[F^{-2} + P^{-2} + (FP)^{-1}]$.

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

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

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 $E(\Delta^2_{F,P})$ require modification.

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

Figure 3 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 the past and the 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. The decrease is 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 50 only slightly decreases the uncertainty about the future sample average.

Another important issue was analyzed in the experiment leading to Figure 4. The top and the 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 toward 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 $H = 0.7$, and $M = 10{,}000$, whereas its long-run properties are nearly identical to those of white noise.

At this point, we wish to digress from dry mathematics and experimental results to explain the specific importance of this last result in hydrology. Statistics, like the rest 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 "always use the simplest model that you can get away with" (because it is not disproved by the data). In this light, let us choose 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 a fit may be measured by the $R(t, \delta)/S(t, \delta)$ statistic, to be discussed in Part 2 of this paper (the next chapter). 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 improves as M increases but ceases to improve if M increases beyond 20. Thus, with short records, a memory of $M = 20$ 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 with $M = 20$.

It would be unreasonable, however, to try to use this best-fit 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 that takes independent values.

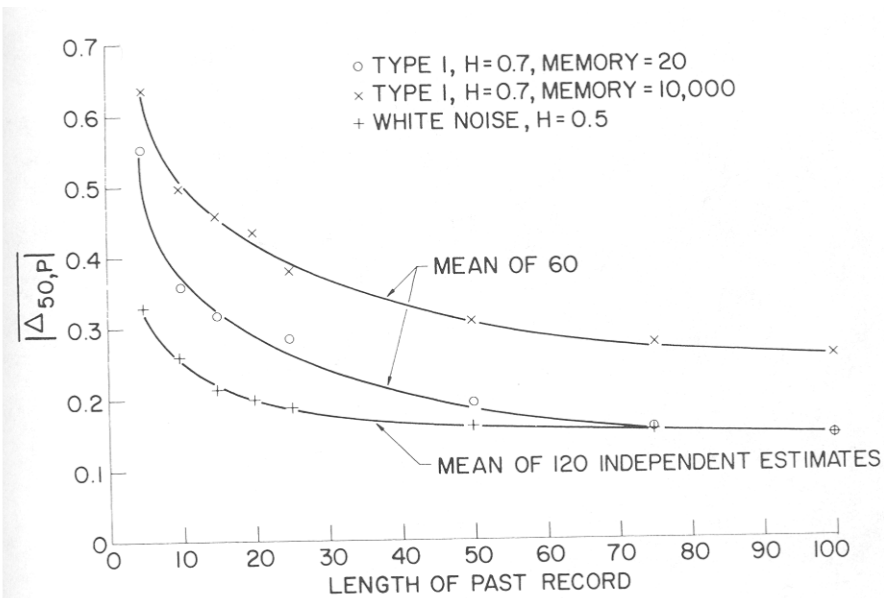


FIGURE C12-3. This figure concerns various Type 1 approximate FGN (as marked). It illustrates the behavior of $\Delta_{50, P}$, defined as the difference between the sample average over the next $F = 50$ years and the sample average of the past P years. The abscissa is P. Our program yielded a sample average Δ_{50}. It would have been easier to compare the mathematical formulas in the text with the behavior of P^2. 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.

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 achieve a reasonable extrapolate from the past P to the future F years, one must consider also effects that are fully developed only when samples of length $P+F$ are available.

To account for such effects, information available in past records is unavoidably deficient, but some additional information may be borrowed from other data believed to resemble the data being studied. In the present context, this general precept can be implemented as follows. Suppose there is evidence that one deals with a stream characterized by $H=0.7$; then 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 for which the available past sample gives a reasonable fit.

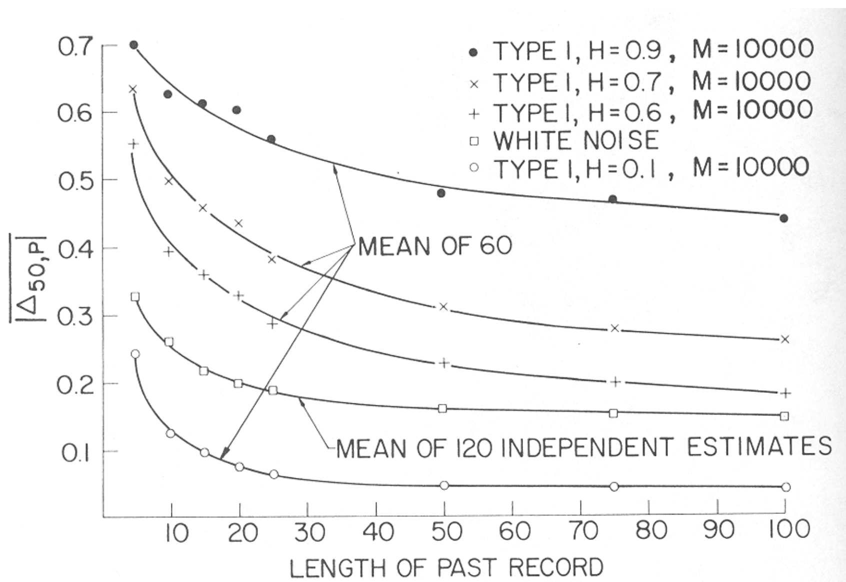


FIGURE C12-4. The effect of several combinations of the memory M and the exponent H upon the behavior of the sample average of $|\Delta_{50,P}|$. (See the caption of the preceding figure for a warning concerning this choice of an average.) Consider a fractional noise with $H=0.7$ and $M=20$ from the viewpoint of $|\Delta_{50,P}|$; we see 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.

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

THE COUNTERPART OF THE 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 practice, one knows neither EX nor EX^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 comparatively complicated, but it will be needed in future applications. In addition it is useful in dispelling the paradox of the least and most erratically fluctuating among fractional noises. However, the reader may proceed directly to Part 2, that is, the next chapter.

In the case of an independent Gaussian random process, the distribution of $\Delta_{F,P}/S^P$ is related to a distribution due to "Student." (His real name was William Gossett, but his employer – Guinness Breweries – forced him to adopt a pseudonym to keep their competitors from knowing that statistics could help brewers make money). the ratio t_p/S_p, where

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

now called "Student's distribution with $P-1$ degrees of freedom." The distribution of the random variable t_P 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

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

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

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

$$S_P^{-1}\left\{P^{-1}\sum_{u=1-P}^{0} X(u) - EX\right\} \text{ and } S_P^{-1}\left\{F^{-1}\sum_{u=1}^{F} X(u) - EX\right\}$$

are no longer independent, and the distribution of their difference $\Delta_{F,P}/S_P$ becomes difficult to derive. To simplify the continuation of this discussion, we shall make the additional assumption that $F = P$ for large P. A heuristic argument suggests the following.

A) $\Delta_{P,P}$ and S_P are nearly independent because $\Delta_{P,P}$ depends primarily upon low frequency effects, whereas S_P depends primarily upon high frequency effects.

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

$$E\left[\frac{\Delta_{P,P}}{S_P}\right]^2 \sim \frac{E(\Delta_{P,P}^2)}{E(S_P^2)} \sim \frac{C_H(4P^{2H-2} - 4^H P^{2H-2})}{C_H(1-P^{2H-2})} = \frac{4(1-4^{H-1})}{P^{2-2H}-1}.$$

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 $t_P(P-1)^{-0.5} - t_F(F-1)^{-0.5}$. For other values of H and P, representative numerical values of V are reported in Table 1. (See page 295).

C) A third prediction of the heuristic argument is that if H is changed from $H = 0.5$ to some other value, $\Delta_{P,P}/S_P$ is unchanged in the form of its distribution but is multiplied by $V(P, H)/V(P, 0.5)$. Using computer simulation, it was established that this 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 $[\Delta_{P,P}/S_P]\,[V(P, 0.5)/V(P, H)]$ is roughly distributed as the difference $t'_P(P-1)^{-0.5} - t''(P-1)^{-0.5}$, where t'_P and t''_P are two independent Student's random variables.

This is a good point to return to the discussion of identifying which are the most "violently fluctuating" among fractional noises. We saw earlier that, for given values of P, C_H, and EX^2, $E[\Delta^2_{F,P}]$ *increases* with H, showing (as expected) that fractional noises with *high* values of H are the *most* violently fluctuating of all. We also saw, however, that S_P *decreases as H increases*. If it was necessary to measure the "violence" of fluctuations by quantities that increase with S_P, our finding that S_P decreases as H increases would have been paradoxical. This implies only that 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. Therefore, no paradox is present. Not only does the ratio $\Delta_{F,P}/S_P$ *increase* with H, but such an increase results from two *combined* effects: 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 $P \to \infty$, it is known that t_P tends towards a Gaussian of zero mean and unit variance so that the limit of $t_P(P-1)^{0.5} - t_F(F-1)^{-0.5}$ is 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" (for example, 5% or 10%), except in a small percentage (for example, 1%) of the cases. Let us inspect the table of Student's distribution; except for 1% of all cases, one finds that t_P is within 10% of its Gaussian limit whenever $P > 20$ and 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 probability p in the case where $H = 0.5$. If $H \neq 0.5$, the value obtained above is simply to be multiplied by V.

For the sake of mathematical precision, the above statement concerning nearly Gaussian distributions may be re-expressed as follows. Let $g(p)$ denote the threshold exceeded with 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 $P > 20$. To ensure that $x_P(p)/g(p)$ lies between 1 and 1.1 for all p between 0.0005 and 0.5, it suffices that $P > 20$. 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 $P > 40$.