Noah, Joseph, and Operational Hydrology

BENOIT B. MANDELBROT

JAMES R. WALLIS

International Business Machines Research Center
Yorktown Heights, New York 10598

COMMENT AND ERRATA

The papers which the present work introduces and summarizes have been postponed to the February and April issues of WRR. In those papers, the complex interplay between the random variable R(s)/S(s), its sample distribution, its expectation and its variance, is described in detail. Therefore, the present work postpones all qualification and handles R(s)/S(s) quite casually. We hope it will be clear to the reader that, depending upon the context, the same letters R(s)/S(s) stand, either for this random variable itself, or for its expectation on or upon a sample estimate of such expectation.

In the following errata, the letters "L" and "R" stand for "left" and "right" column, respectively, "line -13" stands for "line 13 from the bottom of the column".

p. 911 L, line 12, replace 0.5 by $s^{0.5}$

p. 913 R, line -13, replace $\exp(-|s|/s)$ by $\exp(-|s|/s_2)$

p. 913 R, line -11, replace $(1+|s|{}_2/s_3)^{-2}$ by $(1+|s|/s_3)^{-2}$

p. 914 L, line -14, replace ΔX^* by $\Delta X^* =$

p. 915 L, line -3, replace e:f by 1:f

p. 916 L, line 28, replace $C\sqrt{s}$ by $K\sqrt{s}$, with K some (positive and finite) constant.

p. 916 L, line -5, replace X(u) by C(u)

p. 916 R, line 16, replace C by G

p. 916 R, line 17, replace C a constant by G a random variable independent of t and s

p. 916 R, line -8, replace by

$$C_H(s) = Q\left[(s-1)^{2H} - 2s^{2H} + (s+1)^{2H}\right],$$ with Q any (positive and finite) constant.

p. 916 R, line -6, replace by

$$C_H(s) = \left[2H(2H-1)Q\right] s^{2H-2}$$

p. 916 R, line -9, replace $s \geq$ by $s \geq 1$

Water Resources Research: 4, 1968, 909-918. **H10**

Noah, Joseph, and operational hydrology (M & Wallis 1968)

... were all the fountains of the great deep broken up, and the windows of heaven were opened. And the rain was upon the earth forty days and forty nights. *Genesis:* **6**, 11-12.

...there came seven years of great plenty throughout the land of Egypt. And there shall arise after them seven years of famine... *Genesis:* **41**, 29-30.

Dedicated to Harold Edwin Hurst

✦ **Abstract.** This is an introduction to a series of studies concerned with two about precipitation. We propose the terms "Noah Effect" for the observation that extreme precipitation can be very extreme indeed, and "Joseph Effect" for the finding that a long period of unusually high or low precipitation can be extremely long. While both effects are extremely well-established current models of statistical hydrology cannot account for either effect and must therefore be superseded. As a replacement, the "self-affine" models that we propose appear very promising. They account particularly well for a remarkable empirical observation due to Harold Edwin Hurst. ✦

A SERIES OF PAPERS to which this work serves as introduction and summary shall describe in detail a family of statistical models of hydrology which we believe adequately account for the Noah and Joseph Effects. Different papers in the series will be devoted, respectively, to mathematical considerations, to accounts of computer simulations, and to analyses of empirical records.

INTRODUCTION

The models to be described were advanced in M 1965h{H9} and M & Van Ness 1968 {H11}. We have extensively tested and developed them and believe that we have confirmed their soundness. It may be of interest to note that they are instances of a broad family of "self-affine models." The concept of "self-affinity" (without that word) originated in the theory of turbulence, to which it was long restricted, but has recently proved its value in the study of a variety of natural phenomena (see, for example, M 1963b{E14}, M 1966b{E19}, M 1967i{N9} and M 1967s).

A word of acknowledgment is appropriate before we proceed. In investigations of current statistical models of hydrology, one of the most active groups has been that founded by Professor Harold A. Thomas, Jr., at Harvard. Since much that follows is critical of current hydrology, the authors hasten to express their personal indebtedness to Harold Thomas. He directed BBM to Hurst's work and to hydrology, and later initiated JRW to the intricacies of "synthetic hydrology" and simulation.

PARTISAN COMMENTS ON STATISTICAL HYDROLOGY

Current models of hydrology assume precipitation to be random and Gaussian (i.e., to follow the normal probability distribution, with its "bell curve" (often called "Galton ogive"), with successive years' precipitations either mutually independent or with a short-memory. "Independence" means that a large precipitation in one year has no "after effect" on the following years. "Short memory" means that all aftereffects die out within a few years. The classical short-memory mechanism, the Gauss-Markov process, is a "single lag linear autoregressive model." In this case, aftereffects die out in a geometric progression and decrease rapidly. Of greater generality are the "multiple lag linear autoregressive models."

One feature common to all these models is that they belong to "the Brownian domain of attraction"; this is a fundamental notion that we shall define later. It is our belief that models in the Brownian domain cannot account for either the Noah or the Joseph Effects. These models underestimate the complication of hydrological fluctuations. That is, the task of "controlling them by establishing reserves to make the future less irregular" (to paraphrase the title of Massé 1946) is more difficult than these models suggest.

Many models that hydrologists find disappointing (for example, Yevjevich 1968) happen to belong to the Brownian domain. Therefore, the significant feature of our sweeping assertion resides in our blanket condemnation of all models that belong to the Brownian domain. Since this condemnation may appear controversial, let us sketch various "patched up" Gauss-Markov models that have been proposed for the Joseph Effect. Our point will be that, in effect, such models are contrived to behave "as if" they actually were outside the Brownian domain.

Some authors, unhappy with the Gauss-Markov model, nevertheless believe that a description of hydrological reality can be obtained using a Gauss-Markov process whose parameters slowly vary in time. Such models must, however, be changed before their consequences have had time to develop fully. For example, one assumes that, before the sample average of precipitation has had time to "stabilize" near its expected value, climatic change has modified that expectation. We view such models as rather pointless because the usefulness of statistical models lies primarily in their predictions. Since a changing expected value easily overwhelms Gauss-Markov fluctuations, a Gauss-Markov hydrological model cannot be used by itself but only in conjunction with some "master model" that rules climatic change. The overall model, combining hydrology and climatology, is far from being of Gauss-Markov form.

Other approaches to hydrological modeling also start with a Gauss-Markov process, and then introduce modifications that tend to be more extensive when records are long than when they are short. This is illustrated using two examples. The first involves the loose but intuitive idea of the duration of a drought; the second involves the more rigorous but less intuitive concept of the Hurst range.

If an independent Gauss or Gauss-Markov process is chosen to fit precipitation, the duration of the longest drought will be greatly underestimated. Therefore, such processes must be modified to allow for more durable aftereffects (for example, through "multiple lag" models). One who considers such modifications as nuisance corrections to a basic Gauss-Markov process will naturally try to fit all available data with a "minimal" modified process, having as short a span of after-effects as possible. However, as the sample duration increases, "unexpectedly" long droughts are bound to appear. Their appearance shows (after the fact) that the "minimal" model had attributed a special significance to the longest sample T that was available when the model was constructed. As the sample increases, this model must be changed; for example, by increasing the number of lags.

"Drought" being an elusive concept, let us proceed instead to the observed behavior of the "Hurst range," which is less intuitive but easier to study. {P.S. 1999. The definition is illustrated in }. One begins by evaluating the total capacity $R(\delta)$ that a reservoir must have had in order to perform "ideally" for δ years. Here "ideal performance" means that (a) the outflow is uniform, (b) the reservoir ends the period as full as it began, (c) the dam never overflows, and (d) the capacity is the smallest compatible with (a), (b) and (c). The concept of an ideal dam is of course purely retrospective, since data necessary to design such a dam are only known after the fact. However, the past dependence of the ideal capacity upon δ says a great deal about the long-run behavior of a river on which an actual dam is to be built.

Postponing qualifications to later papers, {P.S. 1999, by the same authors, all reprinted in this book}, let us describe a striking discovery that H. E. Hurst made while examining the yearly discharges of the River Nile and other geophysical records. In Hurst 1951, 1956, the ideal reservoir capacity $R(\delta)$ was divided by the standard deviation $S(\delta)$ of δ successive discharges. The empirical finding is as follows:

Except perhaps for small values of δ*, the rescaled range* $R(\delta)/S(\delta)$ *is proportional to* δ^H*, where H is a constant between* 0.5 *and* 1.

Hurst judged H to be "typically" near 0.7, but estimates (to be reported in M & Wallis 1969b{H27}) put H much higher and above 0.85. This is significant because an independent Gauss model yields $R(\delta)/S(\delta) \sim \delta^{0.5}$. Gauss-Markov models, "multiple lag" models and all other models in the Brownian domain give a more complex prediction. They predict that $R(\delta)/S(\delta) \sim \delta^{0.5}$ for large δ, but that $R(\delta)/S(\delta)$ grows faster than $\delta^{0.5}$ for small or moderate δ, which we shall call the "initial transient."

In this transient region, a variety of different behaviors may be obtained. Moreover, many models may lead to the same transient behavior, which makes them indistinguishable from the viewpoint of predictions concerning $R(\delta)/S(\delta)$. Thus, if $R(\delta)/S(\delta)$ is only available for $1 \leq \delta \leq T$ (with T a finite duration), many different models of the Brownian domain are likely to yield predictions indistinguishable from the data. However, for $\delta > T$, the Hurst range of every one of these processes will soon merge into the classical $\delta^{0.5}$ pattern. So far, such a convergence has never been observed in hydrology. Thus, those who consider Hurst's effect to be transient implicitly attach an undeserved importance to the value of T, which in most cases is the largest currently available sample size. These scholars condemn themselves to never witness the full asymptotic development of the models they postulate.

TOWARDS A CHANGE OF DIRECTION IN HYDROLOGICAL MODELING

Our criticism of hydrological models obtained by "patching up" the Gauss-Markov process will be further developed in the sequel. But it should be understood that we expect to put forward an alternative model that involves few parameters yet manages to represent fully the tremendously complicated hydrological reality. A model having few parameters can only be a "first approximation." But the goal of such a first approximation must be to "capture" the main features of the problem, namely the Joseph and/or Noah Effects.

To characterize our proposed non-Brownian first approximations, a loose distinction between "low-frequency" and "high-frequency" phenomena is useful. Using a Gauss-Markov process implies fitting high frequency effects first and worrying about low frequency effects later. We propose to reverse this order of priorities. Conveniently, the term "low frequency" applies equally well to a rarely occurring phenomenon, and to an oscillating phenomenon with a long wavelength.

The concepts of "low" and "high" frequency are, of course, relative. Natural phenomena cover a continuum. For example, the frequency of one cycle per day is perceived as very low in turbulence theory and very high in hydrology. It is also a fundamental frequency in astronomy, so it may also separate zones in which intrinsically different mechanisms rule the fluctuations of precipitation. The same argument may hold for the wavelength of one year.

Another important wavelength is 50 to 100 years, which we shall refer to as a "lifetime." This is roughly the horizon for which one designs water structures and (coincidentally) this is the length of most hydrological records. This wavelength is of human, not astronomical, origin; it is purely "anthropocentric." Whereas precipitation fluctuations of wavelength near one day or one year may participate in several physical mechanisms, those of wavelength near one lifetime are likely to participate in one mechanism only. Thus, the latter are likely to be simpler.

Now, assume that one seeks an approximation valid over a wide band of frequencies. It may be convenient first to look for a good fit in some narrow frequency band, with the hope that the resulting formula will be applicable over the wide band. Under these circumstances, the band near one lifetime, although purely anthropocentric in its definition, constitutes, in our opinion, a better basis of extrapolation than the band near one year, which owes its meaning to astronomy.

We realize that a stress on low frequencies results in an emphasis on idiosyncrasies. But the purpose of hydrological engineering is to guard against the recurrence of such idiosyncrasies; therefore, one cannot afford to neglect any available information.

Models in the Brownian domain have long been recognized as applicable in many fields of science, beginning of course with the Brownian motion of statistical mechanics. As a result, these models' proponents among hydrologists are often able to identify ready-made answers to the standard problems. Our proposed approach requires more work, but the answers appear to be sufficiently better, making this additional work worthwhile. Moreover, we shall see later in the paper that the concept of self-affinity will bring true simplicity.

"MILD" AND "WILD" PROCESSES

We are now in a position to give the promised characterization of the "Brownian domain of attraction," and of the related meanings for the terms "mild" and "wild," as applied to a time series, that is, a function $X(t)$ of the integer-valued time variable t. We need three results of probability theory, two of which are classical, and all three of which relate to averages of T successive terms of $X(t)$.

The first result says that $X(t)$ satisfies the law of large numbers if for $T \to \infty$ its average tends to a limit $EX(t)$, called the expectation. This law gives theoretical justification to the common practice of taking sample averages as estimates of population expectations.

The second result states that $X(t)$ satisfies the more demanding "central limit theorem" in its original form, which asserts the following: For large T, the distribution of the average becomes approximately Gaussian, with a variance tending to zero as $T \to \infty$. This justifies the common belief that if T is not small, the sample average is likely to be a good estimate of the expectation. A corollary of this is that, for large T, even the largest of the T quantities $T^{-1}X(t)$ contributes negligibly in relation to the contribution of the "future average" $T^{-1}\sum_{t=1}^{T} X(t)$.

The third and final result on the averages of random sequences is much less well-known but is very important to applications. Letting the term "past average"denote the expression $T^{-1}\sum_{t=-T+1}^{0}X(t)$, the third basic result asserts that as $T \to \infty$, the past and future averages are increasingly close to being statistically independent. For example, the variance of the difference between them is double the variance of each. It is unfortunate that this property does not yet have a generally accepted name. If a natural phenomenon satisfies the hypotheses of all three of these mathematical theorems, it is said to belong to "the Brownian domain of attraction." We shall also say that the randomness behind the phenomenon is "mild."

Now consider the fractional noises and the approximations thereto used in M & Wallis 1969c. These processes satisfy both the law of large numbers and the central limit theorem, but they fail to obey the hypotheses of the third theorem. Such phenomena, and all phenomena that fail the last two theorems, or even fail all three theorems, will be called "wild." For example, the average $T^{-1}\sum_{t=1}^{t}X(t)$ may *fail* to tend to any limit. Or it may tend to a Gaussian limit, where "past" and "future" averages fail to become asymptotically independent. This latter circumstance is important for the hydrologists. One reason resides in the coincidental equality between the order of magnitude of most past records and the horizon of most designs (both of which equal one lifetime). A second reason resides in the fact that, true expectations being unknown, planning requires the determination of the difference between the expected mean flow over a future lifetime and the known past average.

It is readily verified that in Gaussian models with a limited memory the variability is "mild." On the contrary, the Noah and Joseph effects not only suggest that hydrological data are "wild" but also express the major two forms of wild behavior. We shall speak of "Joseph-wild" behavior when the wettest decade within a century includes an extraordinary "term" of wet years. We shall speak of "Noah-wild" behavior when a few of the years within the century witness "floods" so major as to affect the average precipitation for periods of many years within which the flood years occurred. Needless to say, a process can be both Joseph- and Noah-wild simultaneously, a complication that we shall face much later. "Pure Joseph-wild" behavior will be said to apply when none of the yearly precipitations during a "wet term," had it stood alone, would have been interpreted as a flood.

In practice, the words "wild" and "mild" should not be construed to suggest a "black-and-white" contrast. Indeed, while the mathematical results in question refer to asymptotic behavior, science always deals with finite horizons. Consider, for example, an infinite (nonrandom) time series

$a(t)$. For the mathematician, the basic distinction is whether the sum $\sum_{t=1}^{\infty} a(t)$ is finite or infinite. For the scientist, the ultimate convergence of $\sum_{t=1}^{\infty} a(t)$ is of little importance, unless $\sum_{t=1}^{T} a(t)$ is already close to its limit. Therefore, the concept of "wild" must be considered as allowing for various degrees of intensity, or for various shades of "greyness."

THE MARGINAL DISTRIBUTION OF THE YEARLY FLOW

We shall now characterize more accurately the idea of a "pure Joseph-wild" process. The basic concept here is the "marginal distribution," which is defined as the distribution of the values of a process, when one disregards chronological order. We believe it reasonable to demand that, when the order of values of a pure Joseph-wild sample is scrambled, one should be left with a smooth process. Thus, the marginal distribution of these values will draw a line between, on the one hand, Noah-wild processes and, on the other hand, processes that are either smooth or pure Joseph-wild.

The paragon of the pure Joseph-wild is a process with a *Gaussian marginal distribution.* To check the applicabilities of the Gaussian marginal distributions, it is useful to plot the data on "probability paper", {P.S. 1999: or its computer equivalents. On that paper, the abscissa is a quantity X being tested and the ordinate is the expression Pr $\{X < x\}$ for the Gaussian, which is called error function. Using those coordinates, large Gaussian samples yield straight lines. As applied to hydrological data, probability paper plots show that it is not exceptional for the marginal distribution to be either nearly Gaussian or highly non-Gaussian.

To stay near the land of Joseph, an example of a nearly Gaussian marginal distribution is provided by the level of the Nile at the Rhoda Gauge near Cairo. A highly non-Gaussian example is the annual discharge from Lake Albert. To find other examples of either behavior, it suffices to thumb through Boulos 1951. Straight line interpolations are quite acceptable in certain cases, poor but perhaps bearable in some other cases and dreadful in still other places. A very familiar example of huge deviations from the Gaussian distribution is provided by the runoffs due to major storms, which may appear on histograms as distant "outliers." Also, high water levels, which would be considered "millennium floods" if one extrapolated the tails of the histogram from the body, occur much more frequently in reality than they should under the Gaussian assumption. {P.S. 1999. Many persons know this inequality through its manifestation in finance discussed in detail in M1997E. Under the Gaussian assumption, "ten sigma events have a probability of a few millionth of a millionth of a millionth of a millionth. In fact, they occur all the time in most Financial records.}

Despite the importance of deviations from the Gaussian, we begin our investigation of the Joseph Effect by Gaussian processes $X(t)$. By definition, the joint distribution of the values of such a process at any finite number of instants is a multivariate Gaussian variable. Such processes will be examined in the next several sections. Near the end of the paper, highly non-Gaussian processes with a Noah Effect will be mentioned. (Processes that are only "locally" Gaussian are studied in M 1969e.)

GAUSSIAN PROCESSES AND THE COVARIANCE

Gaussian processes are known to be fully specified by their covariance function; if $X(t)$ is of zero mean and unit variance, the covariance $C(\delta)$ is the correlation between $X(t)$ and $X(t+\delta)$. (Of course, in the case of Gaussian variables, zero correlation is identical to independence.) We want to use the behavior of $C(\delta)$ to classify a Gaussian process as smooth or Joseph-wild. To accomplishing this, we must distinguish between high-frequency, "short-lag" or "short-run" effects – and low-frequency, "long-lag" or "long-run" effects. Short-lag effects depend upon the values of $C(\delta)$ for a few small values of δ. Long-lag effects depend upon the other values of $C(\delta)$. We shall now examine this dichotomy in four examples. The first example is the process of independent increments, whose covariance $C_1(\delta)$ satisfies $C_1(\delta)=0$ for all $\delta \neq 0$. The second example is the Gauss-Markov process of covariance $C_2(\delta)=\exp(-|\delta|/\delta_2)$. The last two examples are the processes of covariances equal to $C_3(\delta)=(1+|\delta|/\delta_3)^{-2}$ and $C_4(\delta)=(1+|\delta|/\delta_4)^{-0.5}$, respectively, with constant δ_2, δ_3 and δ_4.

The above four covariances differ considerably from each other for large δ, but $C_2(\delta)$, $C_3(\delta)$, and $C_4(\delta)$ are all smooth and monotone for small positive δ. Suppose that the sample duration is short, and that the sample covariance is correspondingly "noisy." Then the graphs of $C_1(\delta)$, $C_2(\delta)$, $C_3(\delta)$, and $C_4(\delta)$ may be undistinguishable not only to the eye but also from the viewpoint of many tests of statistical significance that examine each value of δ singly. That is, such statistical tests are liable to indicate that the differences between the sample covariance and any of the functions $C_1(\delta)$, $C_2(\delta)$, $C_3(\delta)$ and $C_4(\delta)$ are not statistically significant for most δ. The statistician could then conclude that all the data will be acceptably fitted after the short-run data have been fitted. Therefore, the statistician will advise the hydrologist that there is no evidence that the data were not generated by an independent Gaussian process (C_1) or a Gauss-Markov process (C_2), or perhaps some other more involved short-memory process.

This would, however, be a rash conclusion. For example, under the hypothesis that the true covariance is $C_1(\delta)$, one would expect the relative proportion of positive to negative sample covariances to be roughly one. This proportion would be larger under the hypothesis that the true covariance is $C_2(\delta)$, still larger if the covariance is $C_3(\delta)$, and larger yet if the covariance is $C_4(\delta)$. Thus, if statistical criteria geared towards low frequency effects can be developed, it is reasonable to expect them to show the same data to be significantly closer to $C_3(\delta)$ or to $C_4(\delta)$ than to $C_2(\delta)$.

Our need, then, is to enhance long-run properties of a process while eliminating short-run wiggles. The best procedure to accomplish this is to integrate or to use moving averages (fancier averages will not be considered here). Three approaches to long-run effects deserve to be identified.

THE VARIANCE OF CUMULATED FLOWS

The "accumulated flow" since time 1 is defined by $\sum_{u=1}^{t} X(u)$ and will be denoted by $X_\Sigma(t)$. Then, G. I. Taylor's formula (see Friedlander & Topper 1961) can be used to evaluate the variance of the increment

$$X_\Sigma(t+\delta) - X_\Sigma(t) = \sum_{u=t+1}^{\delta} X(u) = X(t+1) + \cdots + X(t+\delta).$$

This variance takes the form

$$\text{Var}\ (\delta) = \delta C(0) + 2\sum_{u=0}^{\delta}(\delta - u)C(u).$$

This expression immediately introduces a basic long-run dichotomy based on the values of $\sum_{u=0}^{\infty} C(u) = \Lambda$.

When $\Lambda < \infty$, Var (δ) = Var $\left[X_\Sigma(t+\delta) - X_\Sigma(t)\right]$ is asymptotically proportional to $\delta\Lambda$, and $X(t)$ is in the Brownian domain of attraction.

When, on the contrary, $\sum_{u=0}^{\infty} C(u)$ diverges sufficiently rapidly, Var (δ) grows faster than proportionally to δ and $X(t)$ is *not* in the Brownian domain of attraction. For example, $C(u) = C_4(u)$ yields Var $(\delta) \sim \delta^{1.5}$, where

~ means "asymptotically proportional to." More generally, let $C(u) \sim u^{2H-2}$ for large u, with $0.5 < H < 1$. Then Var $(\delta) \sim \delta^{2H}$ for large δ.

Incidentally, assuming implicitly that $\Lambda = \sum_{u=0}^{\infty} C(u) < \infty$, G. I. Taylor suggested this infinite sum as a measure of the span of memory or temporal dependence in a time series. This quantity is widely believed to be straightforward, but in fact, can be treacherous.

- It is infinite for $C_4(\delta)$;
- It is finite and easy to estimate in cases like $C_1(\delta)$ or $C_2(\delta)$, where the series $\sum_{u=0}^{\infty} C(u)$ converges rapidly;
- Finally, it is finite but difficult to estimate in cases like $C_3(\delta)$, where the series $\sum_{u=0}^{\infty} C(u)$ converges very slowly.

BRIDGE RANGE, THE JOSEPH EFFECT, AND HURST'S LAWS

Curiously, empirical data about the behavior of Var (X_Σ) in hydrology have been examined only recently. The first expression to be examined was a different measure of the overall behavior of a process, namely $R(\delta)/S(\delta)$. Here, the sequential range $R(\delta)$ was defined earlier to be the capacity of a reservoir capable of performing "ideally" for δ years, and $S(\delta)$ was defined to be the standard deviation of yearly flow for δ years. Among Gaussian processes, the dependence of $R(\delta)/S(\delta)$ upon δ sharply distinguishes smooth from Joseph-wild processes. This distinction is already obvious for Joseph's own example of seven years of drought, for which the ideal reservoir needed to be enormous. If wet and dry years alternate, then ideal reservoir size decreases. We propose to express this reduction numerically.

First, consider the case when $X(t)$ is an independent Gaussian process. When δ is large, it follows that both $R(\delta)$ and $R(\delta)/S(\delta)$ equal $\sqrt{\delta}$ multiplied by some "universal" random variable independent of δ. The little information that is known about those random variables is found in Feller 1951. For the Gauss-Markov process and for other models for which the memory $\Lambda = \sum_{u=0}^{\infty} C(u)$ is finite, the "$\sqrt{\delta}$ law" remains true, but the multiplying random variables are no longer the same.

The case of time series exhibiting the Joseph Effect. Their behavior is entirely different: the $\sqrt{\delta}$ law *fails*, as first noted in Hurst 1951, 1956 and Hurst et al 1965. For hydrological series, as well as for many other natural time series, $R(\delta)/S(\delta)$ increases like $C\delta^H$. Here, C and H are positive constants; H may range between 0 and 1 and is seldom near 0.5. We shall call this empirical finding "Hurst's law." Moreover, $\sqrt{\mathrm{Var}[\Delta X_\Sigma]}$ is also proportional to δ^H rather than to $\delta^{0.5}$, as suggested by the usual simple models. This will be called "Hurst's law for the standard deviation," or "Langbein's corollary of Hurst's law" because it was first noted in Langbein's comments of Hurst 1956.

Strictly speaking, Hurst claimed a more demanding "one parameter δ^H law," $R(\delta)/S(\delta) \sim (\delta/2)^H$. His reasons for claiming that $C = 2^{-H}$ are unclear and not convincing. Moreover, it is obvious that separate selection of H and C ensures a better fit; it also yields a different estimate of H. For example, Ven Te Chow, in his contribution to the discussion of Hurst 1951, found a case where H changes from 0.72 to 0.87 when C is estimated separately. Also, we found cases where the best estimate of H is below 0.5, contradicting Hurst's assertion that $0.5 < H < 1$. See M & Wallis 1969b{H27} for a revised value of H.

Note also that Hurst's "one parameter $(\delta/2)^H$ law" has proved dangerous. In some cases it tempted him, as well as others, to estimate H from a single sample of natural or simulated values of $X(t)$. Such estimates should be discarded. The revised statement that we use means that the estimation of H requires many values of δ and, for every value of δ, a large number of starting points t spread over the total sample of length T.

On the other hand, every specific model of the Joseph Effect, such as the fractional noise (to be described in the sequel), will yield a relationship between C and H, whose conformity with experience will test the value of the model.

SPECTRAL ANALYSIS: PRINCIPLE AND APPLICATION IN HYDROLOGY

In addition to Var (ΔX_Σ) and Hurst's range, a third way of looking at low frequency phenomena is through spectral (or Fourier or harmonic) analysis. The sole reason to mention it here is because the spectral density of hydrological records peaks sharply for very low frequencies, as is also the case for the so-called $1/f$ noises (M 1967i{N9}). A full discussion of this topic will be found in Part 2 of M & Wallis 1969a {H12}.

RELATIONS AMONG THE JOSEPH EFFECT, HURST'S LAW AND THE GLOBAL BEHAVIOR OF THE COVARIANCE

To account for the observed behavior of Var (δ), $R(\delta)/S(\delta)$, and the spectrum has proved to be a very hard task. For example, perusal of the discussion of Hurst 1951 demonstrates the desperate expedients to which he found it necessary to resort, in order to fit his finding within the familiar universe of simple statistical models. Claiming (incorrectly, as we shall demonstrate) that there exists no stationary random process with a range following the δ^H law, several discussants have suggested either abandoning statistical stationarity or invoking nonrandom "climatic" changes.

A more helpful reaction, already mentioned in the partisan comment at the beginning of this paper, is exemplified in Anis & Lloyd 1953 and Fiering 1967. These and other authors have constructed stationary stochastic processes of the usual kind: those in the Brownian domain of attraction, satisfying $\sum_{u=0}^{\infty} C(u) < \infty$ for which both the range and the standard deviation are proportional to δ^H over a finite span of values of δ. But the usual $\sqrt{\delta}$ behavior still applies beyond this span. Thus, the δ^H law is, for these authors, a property of what we have called a transient span. This transient may be made arbitrarily long. But long transients can only be achieved with complicated processes having a long memory. For example, Fiering 1967 (p. 85) had to use an autoregressive model with 20 lags (an exorbitant number of lags) to ensure that Hurst's law holds over the span $1 < \delta < 60$.

An alternative to this approach is based upon the existence of the self-affine random processes (see M 1965h {H9}). For these processes, Hurst's law holds for all values of δ. Even more important from our viewpoint, which emphasizes low frequency phenomena, is the existence of processes for which Hurst's law holds for the short run as well as for the long run. For the standard deviation, this was already proved when we noted in passing that

$$\sqrt{\mathrm{Var}[\Delta X_\Sigma]} \sim \delta^H \text{ holds if and only if } C(\delta) \sim \delta^{2H} - 2.$$

The asymptotic behavior $\sim \delta^H$ *can be shown to hold also for the range* $R(\delta)$ *and the rescaled range* $R(\delta)/S(\delta)$.

This observation is central to our study of the Joseph Effect. Before examining it more closely, let us show how it can explain the existence of models in which Hurst's law holds during the initial transient. The key is that the values of Var $(\delta) = \mathrm{Var}\left[X_\Sigma(t+\delta) - X_\Sigma(t)\right]$ for $\delta < T$ are affected only by the values of $C(\delta)$ for $\delta < T$. Hence, changes in the covariance for $\delta > T$ pass unnoticed when only the span $\delta < T$ is observable. Now suppose that, starting from the covariance $C(\delta) = (1 - 10\delta)^{2H-2}$, long-lag covariances ($\delta > T$) are decreased sufficiently to make $\sum_{u=0}^{\infty} C(u)$ convergent. The result is that the modified process $X(t)$ is "brought back" into the Brownian domain of attraction. It could even be made into a "multiple lag" autoregressive model, which is the usual generalization of the Markov model. For such a modified process, Hurst's law continues to hold for $\delta < T$ and for some time beyond $\delta = T$. In the long run, however, it will be replaced by the $\sqrt{\delta}$ law. For example, the standard deviation will equal $K\sqrt{\delta}$, with K equal to some (positive and finite) constant. The value of this constant depends upon the tail selected for the modified covariance and it is adjustable at will.

We consider the models in which T plays a central role, to be *undesirable*.

DEFINITION OF SELF-AFFINE WILD GAUSSIAN PROCESSES

In criticizing the usual statistical models as applied to hydrology, we do not underestimate their positive features. In particular, whenever an independent Gaussian process is an acceptable approximation, it is unbeatable. If it is not an acceptable approximation, some of its features deserve to be preserved. For example, all models in the Brownian domain preserve the assumptions that the variance and Taylor's scale $\sum_{u=0}^{\infty} C(u)$ are finite. But they destroy another property that makes the independent Gaussian processes particularly convenient to manipulate. We want to express this property in a form that is indirect but is easier to generalize. The independent Gaussian process $X(t)$ is exceptionally convenient because $\sum_{u=1}^{t} X(u)$ can be interpolated to continuous times with the help of a "self-affine" random process $B(t)$, called Brownian motion (also called a Bachelier process, or a Wiener process). To define self-affinity, one must consider a portion of $B(t)$, with t varying from 0 to T, and rewrite it as $B(h, T)$, with h varying from 0 to 1. "Self-affinity" then expresses the fact that the rescaled function $T^{-0.5}B(h, T)$ has the same distribution for every value of T.

From this, one immediately deduces that

$$\sqrt{\mathrm{Var}}\ \left[B(t+\delta) - B(t)\right] = s^{0.5} \text{ and } \max_{0 \le u \le \delta} B(t+u) - \min_{0 \le u \le \delta} B(t+u) = C\delta^{0.5},$$

where C is a random variable independent of t and δ. These statements are forms of the $\delta^{0.5}$ law, but they are valid uniformly (that is, for all δ) rather than asymptotically (that is, for high δ).

By analogy, when studying the laws of Hurst, it is good to know that more general self-affine processes exist. A Gaussian process $X(t)$, of the integral $B_H(t) = \int_0^t X(u)du$, is Joseph self-affine if the rescaled function $T^{-H}B_H(h, T)$ is independent of T in distribution. That the δ^H laws apply to $B_H(t)$ can be seen by simple inspection. $B_H(t)$ was called "fractional Brownian motion" by M & Van Ness 1968 {H11}.

Unfortunately, the derivative $B'_H(t)$, called the "fractional Gaussian noise," is too irregular to be studied directly. Previously, we interpolated the integral of the independent Gauss process by Brownian motion; now, we must replace $X(t)$ by $B_H(t+1) - B_H(t)$. The covariance of the function $B_H(t+1) - B_H(t)$ is given for $\delta \geq 1$ by

$$C_H(\delta) = Q[(\delta - 1)^{2H} - 2s^{2H} + (\delta + 1)^{2H}]$$

where Q is any (positive and finite) constant. If $1/2 < H < 1$ and δ is large, the covariance is approximated by

$$C_H(\delta) = [2H(2H - 1)Q]\delta^{2H-2}.$$

This is precisely the form that we proposed to model phenomena obeying Hurst's law. Clearly, our models are approximations to fractional Gaussian noise.

IMPLICATIONS OF SELF-AFFINITY

To apply self-affinity, one proceeds in a manner similar to the classical "dimensional analysis" of fluid and solid mechanics. This similarity is no accident; hydrology can be considered the low frequency application of the theory of turbulence, into which self-affinity and dimensional analysis were introduced by von Kármán and Kolmogorov (see Friedlander & Topper 1961).

A familiar illustration of dimensional analysis occurs in the study of water flow in vessels having the same shape but different dimensions and proportionately different velocities. The calculations need not be repeated. It suffices to solve all relevant problems once. Solutions to other cases will be obtained by mere rescaling.

The application of self-affinity to hydrology follows in the same spirit. Once one performs the calculations relative to some "reference" horizon, answers relative to other horizons are obtained by simple rescaling. This may make it worthwhile, in the case of the reference horizon, to perform some very lengthy and involved calculations that would not otherwise be economic. The new hydrology we propose may demand readjustments of thought. But there is hope that the ultimate outcome of this new hydrology will be a set of new and better "cookbook recipes."

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

Editorial changes. "Self-affine" is used in this reprint as a replacement for "self-similar." The historical confusion concerning these terms is discussed in Chapter H1. The original paper characterizes certain random process as being either "smooth" or "erratic." Those early and poorly chosen terms have been replaced by "mild" or "wild," respectively, following the lead of M 1997E, Chapter E5. The original referred to several planned publications that did not materialize. Here is a shortened version of the last section of the original.

Foretaste of the Noah Effect. To explain Hurst's finding, the Noah Effect is neither necessary nor sufficient. A discussion of the Noah Effect is several papers removed from the present introductory article, {P.S. 1999. This discussion was never completed, but this paragraph was not erased}. Our approach will resemble the methods M 1963b {E14} which are important but mostly used to describe the variation of commodity prices. Consider the function $X_\Sigma(t) = \sum_{u=1}^{t} X(u)$, with $X(u)$ being the annual flow for the year u, minus a reference level around which $X(u)$ oscillates. This sum will be the counterpart of the price of a commodity at the instant t. The very rapid and large changes typical of the behavior of prices will be compared to floods. Sporadic processes (see M 1967b {N10}) will also be needed.

How my fruitful collaboration with J. R. Wallis came about. This story is told in Section 4.3 of Chapter H8.