Proc. National Academy of Sciences (USA) 72, 1975, 3825-3838 **H19**

Earth's relief, shape and fractal dimension of coastlines, and number-area rule for islands

• *Chapter foreword.* This is the first statement of my simplest fractal theory of relief, which became widely popular. Earth's relief is so familiar to anyone that there is no need here to refresh the reader's memory by providing examples. Today, this paper is mainly of historical interest, but a few points made here have not been restated anywhere else. •

✦ **Abstract.** The degree of irregularity in oceanic coastlines and in vertical sections of the Earth, the distribution of the numbers of islands according to area, and the commonality of global shape between continents and islands, all suggest that the Earth's surface is statistically self-affine. The preferred parameter, one which increases with the degree of irregularity, is the fractal dimension D of the coastline; it is a fraction lying between 1 (limit of a smooth curve) and 2 (limit of a plane-filling curve). A rough Poisson-Brown stochastic model gives a good first approximation of the relief by assuming that it is created by superposing very many, very small cliffs, which are placed along straight faults, and are statistically independent. However, the predicted relative area for the largest island is too small, and the predicted irregularity for the relief is excessive for most applications. The predicted dimension is $D = 1.5$, which is, likewise, excessive. Several higher-approximation self-affine models are described. Any model can be matched to the empirically observed D and can link all observations together. But self-affinity must be postulated and cannot yet be explained fully. ✦

ONE BROAD TASK OF GEOMORPHOLOGY is to differentiate two aspects of relief that are best contrasted using electrical engineering terminology. (a) The first is a "signal," defined as a reasonably clear-cut feature that one hopes to trace to a small number of tectonic, isostatic or erosional causes. (b) The second is a "noise," defined as a feature that one believes is due to many distinct causes that have little chance of being explained or even disentangled. This paper is mainly concerned with the examination of this "noise."

The fit between the models and the empirical relationships that they aim to represent will prove to be good. Its quality may even seem surprising, since there was no objective way to know *a priori* whether the relationships that it seeks to represent are indeed mostly noise-related. In fact, even the simplest model will generate ridges, that one is tempted *a priori* to classify as signal-like. Thus, our study of noise will conclude by probing the intuitive distinction between it and the signal.

The strategy used in this paper is, in part, very familiar, having proved successful in taming the basic electric noises. In moving from noises to relief, the first novel aspect (and a major difficulty) is that it deals not with random curves but with surfaces. The second novel aspect follows as a consequence of the first. Since an acoustic or an electric noise is not visible (except, perhaps, as a drawing on a cathode ray tube), geometric concepts do not enter in its study, except at a late stage and in an abstract fashion. For the Earth's relief, the opposite is true. However, ordinary geometric concepts are hopelessly underpowered here, implying the need for new mathematics. For our purposes, remembering that the Earth is roughly spherical overall would only bring insignificant corrections. Therefore, we shall assume that, overall, the Earth is flat with coordinates x and y.

1. GOALS

Quantified goals

To be fully satisfactory, a model of relief must explain the following rules or concepts, each of which is a theoretical abstraction based on actual observations. To be merely satisfactory, a model of relief must show at least that these rules and concepts are mutually compatible.

(a) Korcak's empirical number-area rule for islands (Fréchet 1941) states the relative number of islands whose area exceeds A is given by the power-law $\Phi(A) \sim A^{-K}$. A fresh examination of the data for the whole Earth yields $K \sim 0.65$. More local (and less reliable) estimates using restricted regions range from 0.5 for Africa (one enormous island and others whose sizes decrease rapidly) up to 0.75 for Indonesia and North America (less overwhelming predominance of the largest islands).

(b) A second concept states that, even though coastlines are curves, their wiggliness is so extreme that it is practically infinite. For example, it is not useful to assume that they have either well-defined tangents (Perrin 1913), or a well-defined finite length (references are given in M 1967s. Specific measures of the length depend on the method of measurement and have no intrinsic meaning. For example, let a pair of dividers "walk" along the coast; as the step length G is decreased, the number $N(G)$ of steps necessary to cover the coast increases faster than $1/G$. Hence, the total distance covered, $L(G) = GN(G)$, increases without bound.

(c) Richardson's empirical rule asserts that $N(G) \sim G^{-D}$. In this relation, the exponent D is definitely above 1 and below 2; it varies from coast to coast, a typical value being $D \sim 1.3$ (references are given in M 1967s).

(d) M 1967s proposes that it is useful, in practice, to split the concept of the dimension of a coastline into several distinct aspects. Being a curve, it has the topological dimension 1, but the behavior of $L(G)$ suggests that from a metric viewpoint, it also has a "fractal dimension" equal to Richardson's D. Curves of fractional (Hausdorff) dimension have been known for over half a century as an esoteric concept in pure mathematics, but M 1967s has injected them into geomorphology. The notion of a fractal dimension has also found applications in several other empirical sciences (see M 1975f{H18} and M 1975O). Implicit in M 1967s was the further concept that the surface of the Earth has the dimension $D+1$, which is constrained between 2 and 3.

(e) Vertical cuts of the Earth have been studied less thoroughly than the coastlines, which are horizontal cuts, but a model of the Earth should embody all that is known on their behalf (Balmino, Lambeck & Kaula 1973).

Subjective goals: resemblance in external appearance

Another property to be explained by a model of relief is obvious but not easy to quantify. Roughly speaking, it is difficult to distinguish between small and big islands, unless one either recognizes them or can read the scales. One possible explanation argues that the determinants of overall shape are *scaleless,* and hence are not signals but rather noises. In fact, many islands look very much like distorted forms of whole continents, which is perhaps too convenient to be true. Why should self-affinity extend to the tectonic plates?

Thanks to advances in computer simulation and graphics, there is a new test of the validity of a stochastic model. It need no longer be tested solely through the quality of fit between predicted and observed values of a few exponents. It is my belief (although perhaps a controversial one) that the degree of resemblance between massive simulations and actual maps or aerial views must be accepted as part of scientific evidence.

2. STRATEGY

First strategy: an explicit mechanism, its limit behavior and the concept of self-affinity

Our first strategy is the one used to explain thermal noise in electric conductors through the intermediary of shot noise, which is the sum of the mutually independent effects of many individual electrons. The analogous "Poisson-Brown" primary model of the Earth has isotropic increments. It is simple, explicit, direct and intuitive, and, in general terms, fulfills the goals that were listed. In particular, it predicts that coastlines are not rectifiable, that the above power laws are valid and that D must be greater than 1. In fact, the relief that it yields is self-affine.

This last concept quantifies the first subjective goal listed above. It indicates that the altitude $Z(x, y)$ has a property of spatial homogeneity that can be expressed as follows. Select the origin of coordinates so that $Z(0, 0) = 0$, and pick an arbitrary rescaling factor $h > 0$. Self-affinity postulates that $Z(hx, hy)$ is identical in statistical properties to the product of $Z(x, y)$ and some factor $f(h)$. Unfortunately, the predicted value $D = 1.5$ is not satisfactory. No single value can represent the relief everywhere, and most observed values are well below 1.5. This discrepancy is confirmed by the excessively irregular appearance of the simulated primary relief and coastlines.

Second strategy: self-affinity and the use of limits

The Poisson-Brown model must be improved upon, for both numerical and perceptual reasons. To do so, one must unfortunately resort to a strategy that is indirect, more complex, less powerful and less convincing.

First postulate. Without attempting to describe any specific mechanism, we preserve the assumption that the noise element in the relief is the sum of many independent contributions. It follows, for example, that the increment $Z(P') - Z(P'')$ between the two points P' and P'', must belong to a very restricted family of random variables. In addition to the Gaussian, this family includes other members (so far, ill-known in the applications; see below).

Alternative postulate. We adopt the self-affinity of the relief as an excellent quantification of our first subjective goal and, hence, as a summary of many quantified goals. Even after self-affinity is assumed, the identification of an appropriate $Z(x, y)$ continues to pose a challenge. We shall examine several possibilities, discussing them in order of increasing complication, and shall show that they fulfill our various goals in a way that relates them to each other.

3. POISSON AND BROWN SURFACES

Spatial construction patterned on Poisson shot noise

Start with an Earth of zero altitude, then break it along a succession of straight faults and displace the relief vertically on one side of each fault to form a cliff. At this stage, the terms "fault" and "cliff" are to be understood in purely geometrical terms, with no tectonic implication. The resulting relief is denoted by $\prod(x, y)$. It is convenient to choose an origin $(0, 0)$ and to maintain $\prod(0,0) = 0$, but changing the origin only adds a constant to $\prod(x,y)$. (This model obviously neglects the basic roles of isostasy and of erosion.)

The positions of the faults and the heights of the cliffs are assumed random and mutually independent, the former being isotropic with a high average density and the latter having zero mean and finite variance (implying that large values are very rare). A computer simulation is exhibited in Figure 1, showing a perspective view of the relief, and in Figure 3B, showing a larger piece of coastline. The primary relief has already been described as self-affine with $D = 1.5$. Let us now examine its quantitative properties individually.

Vertical cuts

Isotropically random and mutually independent faults have the following property, which is useful to characterize them. Their points of intersection with a straight line (parameterized by u) form a Poisson point process $\prod(u)$. The angles of intersection are distributed uniformly between 0 and 2π. Denote the average number of points of intersection per unit length by λ. Each primary vertical cut is a Poisson random walk. It differs from an ordinary random walk because the instants when it moves up or down are not uniform in time but follow a Poisson distribution.

Limit vertical cuts

Divide $\Pi(u)$ by $\lambda^{1/2}$, which rescales the cliff heights to make them decrease as their number increases; then let $\lambda \to \infty$. As is well-known, the distribution of the Poisson steps in this limit becomes increasingly irrelevant. By the central limit theorem, $\Pi(u)\lambda^{-1/2}$ tends to a Brownian motion $B(u)$.

This limit is a continuous process, meaning that even the highest contributing cliff contributes negligibly to the sum. The overall resemblance between Brownian motion and real vertical sections was pointed out in M 1963e{E3 } (p. 435). It is likely that other authors have noted it earlier, but I have not found any references to that effect. This result is confirmed by spectral analysis (Balmino, Lambeck & Kaula 1973).

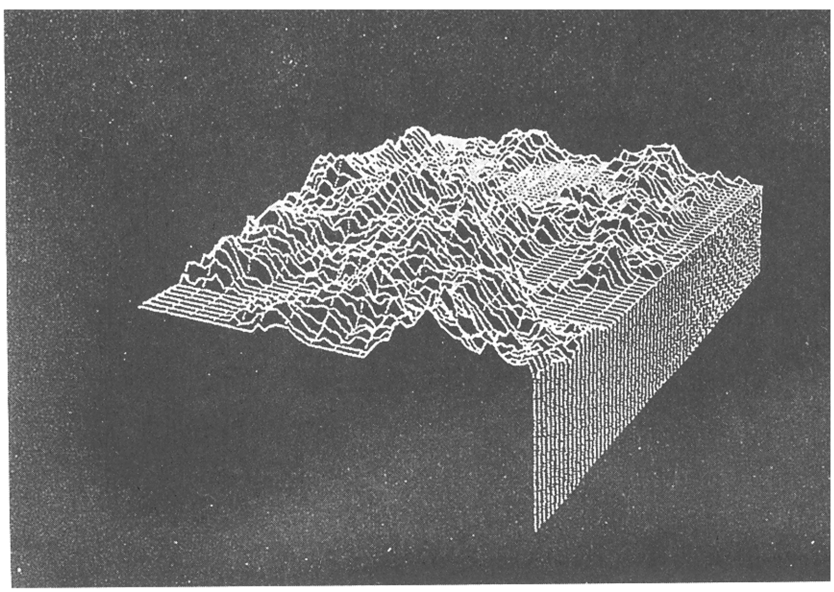

FIGURE C19-1. Perspective view of a sample of a Brownian surface of Paul Lévy. A flat surface at sea level is included to enhance the detail.

Relief

$Z = \prod(x, y)$ can be called a Poisson surface, and $Z = B(x,y)$ is called a Brownian surface by mathematicians. It was defined in Lévy 1965 through the characteristic property that, for every two points P' and P'', $B(P') - B(P'')$ is a zero-mean, Gaussian random variable of variance $|P'P''|^{2H}$, with $H = 0.5$. $B(x, y)$ is self-affine with the scaling factor $f(h) = h^{1/2}$. Until the present application, it was known simply as a mathematical curiosity. Its use as a model could have been introduced directly and dogmatically as just another instance of the oft-successful tactic, which approaches every new statistical problem by trying to solve it using the simplest Gaussian process. However, the detour through Poisson faults improves the motivation.

Coastlines

An island is defined as a maximal connected domain of positive altitude. A coastline, being simply a horizontal section of the relief, has the same degree of irregularity as a vertical section. The coastline of a Brownian island has infinite length, however small its area A; for a Poisson island of area $A \gg \lambda^{-2}$, one has $N(G) \sim G^{-1/2}\lambda^{1/2}$. In either case, when $G \gg \lambda^{-2}$, one has $N(G)^{-1.5}$.

The number-area rule for islands

For islands defined through $B(x, y) \geq 0$, one has $\Phi(A) = A^{-3/4}$ for all values of A. For islands defined through $\prod(x, y) \geq 0$, one has $\Phi(A) = A^{-3/4}$ for $A \gg \lambda^{-2}$.

4. ANISOTROPIC STRETCHING AND ADDITION OF SPECTRAL LINES

A striking feature of sample Brownian surfaces (see Figure 1) is the invariable presence of clear-cut ridges. While they are merely an unexpected consequence of continuity, their presence expresses that each sample is grossly non-isotropic. Since these ridges have no privileged direction, they are quite compatible with the isotropy of the mechanism by which $B(x, y)$ is generated. If we did not know that they are expressions of noise, we might say that they are signals. That is, if we did not know that they are due to the superposition of many effects, we might try to explain them by some single cause.

Nevertheless, they are far from being as regular as either the Appalachians or the Andes, which are profoundly nonisotropic. We shall list two simple accounts for their occurrence by invoking "signals" superposed on a primary "noise."

First signal. This approach consists in a controlled degree of anisotropy introduced into either $\prod(x, y)$ or $B(x, y)$. One may, for example, make the probability of faults greater along some direction than along its perpendicular direction. Alternatively, one can stretch the plane. Either change will make our ridges tend to become parallel to each other and to form mountain ranges; by adjusting the degree of stretching, the overall fit can be improved. In both cases, the values of the Korcak parameter K and the Richardson parameter D would remain unchanged.

Additional signal. A different approach to the problem of non-isotropy is best explained in spectral terms. The spectrum of $B(P)$ is continuous, with a density proportional to ω^{-2}. Just as in communication technology, the signal may be assumed to take the form of a pure spectral line. This assumption would induce sinusoidal up- and down-swells in the relief, thus creating a tendency towards parallel ridges.

V. FRACTIONAL BROWNIAN RELIEF

The most satisfactory model, among those currently available, combines either of the above signals with the noise we are now going to define.

A Gaussian secondary model with either $1 < D < 1.5$ or $1.5 < D < 2$

We now adopt a second strategy and assume that $Z(x, y)$ is a Gaussian surface, meaning that for any set of points $P_n (0 \leq n \leq N)$, the N-dimensional vector of coordinates $Z(x_n, y_n) - Z(x_0, y_0)$ is Gaussian. The combination of isotropy of the increments with self-affinity requires $Z(x, y)$ to be proportional to either the above Brownian function $B(x, y)$ or a generalization, which I propose to call a fractional Brownian function and to denote by $B_H(P)$.

This generalization is defined by $E\{B_H(P') - B_H(P'')\}^2 = |P'P''|^{2H}$, with $0 < H < 1/2$ or $1/2 < H < 1$. In the special case $H = 1/2$, one has $B_H(P) = B(P)$. The fractional Brownian function of time has served the author in modeling a variety of natural time series (M & Van Ness 1968{H11}). The present multiparameter $B_H(P)$ has been mentioned fleetingly in the literature (references are given in M 1975O), but here it

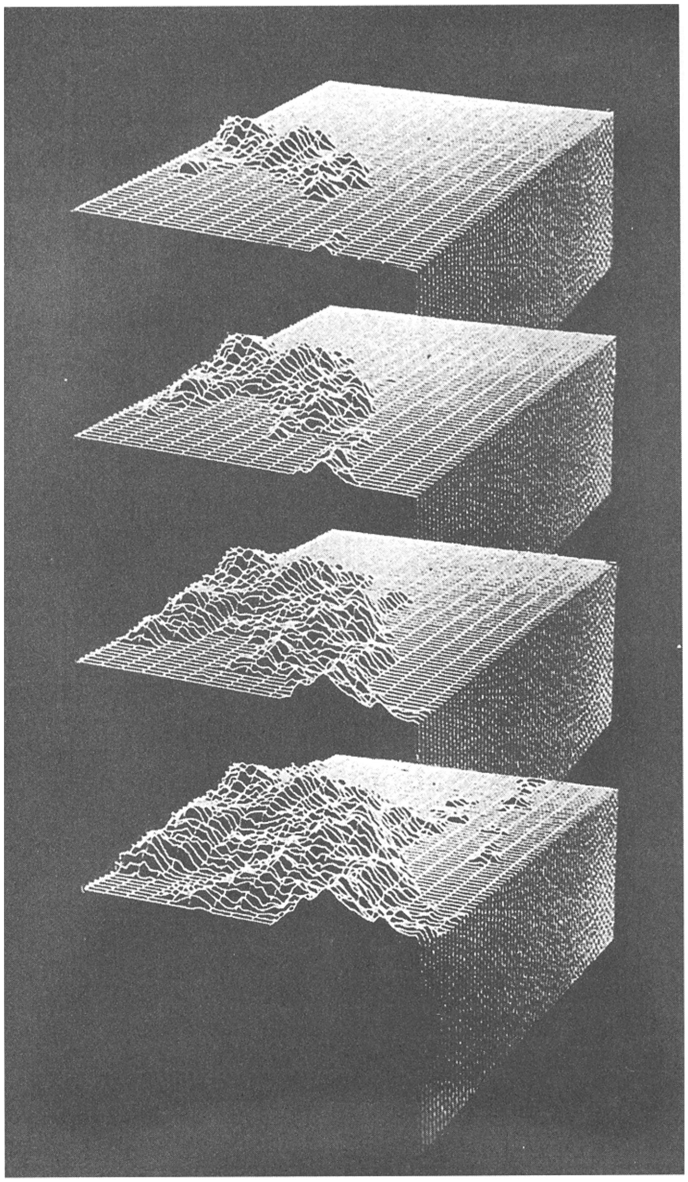

FIGURE C19-2. Several perspective views of a sample fractional Brownian surface for $H = 0.7$ drawn using the same random generator seed as used in Figure 1. Letting the sea level recede, further enhances the shape of the relief. The value of $H = 0.7$ used in this figure gives the best fit from all viewpoints.

can be applied. Selected simulations are illustrated in Figures 2, 3a, and 3c. To satisfy our numerical goals with any desired D, it suffices to select $H = 2 - D$. If $D \sim 1.3$ and $H \sim 0.7$, the subjective goal of familiarity of appearance is fulfilled also.

We end with the Korcak exponent K. Its theoretical value is $K = D/2$ (meaning that the distribution of the typical length $A^{1/2}$ is hyperbolic with the exponent D). The single Earth-wide estimate $K \sim 0.65$ is a compromise between different regions, and indeed fits the world-wide compromise $D \sim 1.3$. Local estimates for Africa and Indonesia also fit the local estimates of D, and the empirical relationship between D and K seems to be monotonically increasing. This feature, if confirmed, would provide an unexpected link between the local and global properties of the relief in different areas of the Earth. This is an interesting topic for further study.

After-the-fact partial rationalization of $B_H(P)$

There are at least two approaches to rationalize $B_H(P)$ each of which is more reasonable in different regions of the Earth.

First method. Note that the spectrum of $B_H(P)$ is continuous with a density proportional to $\omega^{-(2H+1)}$. When $H > 0.5$, it differs from the ω^{-2} spectral density of $B(P)$ by being stronger in low frequencies and weaker in high frequencies. The replacement of $B(P)$ by $B_H(P)$ could be viewed as due to another signal. Its overall effect tends to smooth; its local aspects could easily be associated with erosion, while its global aspects may relate

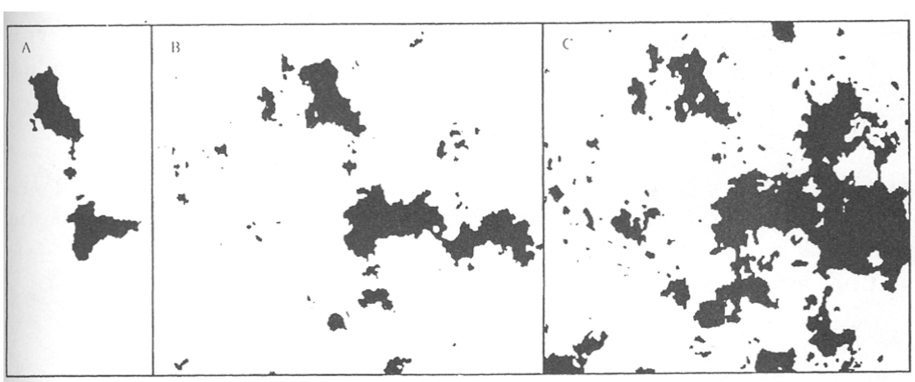

FIGURE C19-3. Several coastlines defined as the zero level lines of fractional Brownian surfaces. They correspond to the same random generator seed but different values of H, namely, $H = 0.7$, $H = 0.5$ (a Brownian surface), and $H = 0.1$.

to isostasy. The fact that D varies around the globe, a fact for which we have not yet accounted, would result from local variability in the intensity of such erosion. However, none of the common methods of smoothing, such as local averaging, is sufficiently effective because each affects only a narrow band of frequencies. The smoothing required here must involve a very broad band; therefore, it would necessarily combine a wide collection of different narrow-band operations. In addition, their relative importance should take a very specific form.

Alternative method. One may obtain $B_H(P)$ directly (this approach is used and is described in M 1975f{H18}) by resorting to cliffs with a special kind of profile. They must rise very gradually but do so forever on both sides of each fault.

6. A NON-GAUSSIAN SECONDARY MODEL

$B_H(P)$ gives a surprisingly good phenomenological description of the relief. But its continuity implies that it will not fit some data. However, discontinuity, if required, is within easy grasp. It suffices to proceed to random surfaces that follow one of the non-Gaussian distributions that may apply to sums of many independent addends, namely, a stable distribution of Paul Lévy. They can be injected into the Poisson model (see above) if cliff heights are assumed to have an infinite variance and to fulfill other requirements. The resulting Poisson-Lévy model cannot be described here. It suffices to say that the largest contributing cliff no longer becomes relatively negligible as the number of contributions increases, but continues to stand out. Without question, we will be tempted to interpret it as a signal.

7. GENERALIZATIONS

The above models are closely related to some recent work in turbulence and have many other fairly immediate applications. They are readily translated to account for such phenomena as the distribution of minerals and oil.

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

A footnote in the original text. A referee, expressing an opinion that may be shared by other readers, criticized my approach for "a narrow focus on purely tectonic processes of the simplest kind and a belief that (i) it is both good and important to make stochastic models whose realizations agree with the largest scale behavior and (ii) if this can be done, it is right and wise to think of these largest scale phenomena as in fact stochastic." To respond, I plead guilty: while the validity of models of the kind that I shall describe must eventually be discussed with great care and skepticism, I see only benefits in first developing them in some detail.

How this paper came to be written. This story is told in Section 4.5 of Chapter H8.