

Unsaturated Zone Hydrology for Scientists and Engineers

James A. Tindall, Ph.D.

United States Geological Survey, National Research Program
Department of Geography and Environmental Sciences, University of Colorado Denver

James R. Kunkel, Ph.D., P.E.

Knight Piésold, LLC, Denver, Colorado;
Department of Geology and Geological Engineering, Colorado School of Mines

with

Dean E. Anderson, Ph.D.

United States Geological Survey, National Research Program



PRENTICE HALL
Upper Saddle River, New Jersey 07458

Spatial Variability, Scaling, and Fractals

INTRODUCTION

In the previous chapters, traditional methods for investigating the unsaturated zone have been discussed: the importance of physical properties; microscopic parameters such as the double-layer theory; behavior of clays; water flow; gaseous diffusion; contaminant transport, as well as other parameters. The concepts discussed in this chapter deal with treating and analyzing the data gathered from unsaturated-zone studies. This discussion is meant to introduce the reader to the basic concepts involved with some of the tools available to us for treating data, while incorporating the various methods used. Because of the importance of spatial variability and understanding its importance in soils research, the reader is urged to study other texts that discuss such material in detail; the same is true for geostatistics and fractals.

Geostatistics has been used for several decades, and precedes the advent of both scaling and fractals. Indeed, it can be said that scaling begins where geostatistics ends; fractals are somewhat similar to scaling, but are much more complicated. The advantage of using geostatistics is the intuitive nature of the process, coupled with the fact that they provide a better tool for analysis than simple statistics. Scaling and fractals are not as intuitive and therefore, are more difficult to comprehend. However, they are additional tools in our arsenal for problem solving and as such, can be quite useful in various investigations of soils research. The concepts discussed here aim at whetting the reader's appetite to seek further knowledge concerning the use of geostatistics, scaling, and fractals.

16.1 FREQUENCY DISTRIBUTIONS OF SOILS

As soils developed throughout eons of time, they became the product of the very factors that helped to form them. These factors included such parameters as climate; parent material; topography; microbial organisms; and time. Soils are heterogenous rather than homogeneous, simply because of the variability in their formation processes due to freezing, thawing, shrinking, swelling, and so on (see chapter 2). The optimization of environmental resources and resource allocation make it necessary to quantify soil spatial variability, and to determine the scale of its occurrences. The need to generate explanations of observable variability and predictability through modeling efforts is addressed here. There are two categories of variability for most landforms: systematic, and random. The classification of a given medium into one of these categories is determined by the number of observations made. As more observations are made, the existing variability naturally decreases in importance, and confidence limits are smaller. In a strict statistical sense, the more the data distribution differs

from a normal distribution, the larger the sample size has to be, for an adequate approximation. However, the cost of collecting large amounts of data is prohibitive and the number of samples needed is greatly influenced by spatial variability. Thus, analytical expressions that require simple solutions with a minimum number of samples are desired.

Because soils are variable, it is important to obtain an estimate of error associated with the specific parameter we wish to measure. This is done with statistics, in which—for simple estimates—the mean (average value of the parameter) and standard deviation (range of the parameter) can be obtained easily. Mathematically, an estimate of the mean is determined by

$$m = \frac{\sum x_i}{n} \quad (16.1)$$

where m is the mean, x_i is measured values for the parameter, and n is the sample number. An estimate for the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum (x_i - m)^2}{(n - 1)}} \quad (16.2)$$

where σ is the standard deviation. For a homogeneous soil, σ values between samples are likely small, indicating that the properties of each of the collected samples are similar. However, if σ values are large, the samples are likely dissimilar. This refers more to variance than the mean; the mean can vary little between media types. An example of variance change is the comparison of bulk-density values for sandy-textured soil versus a clay-textured soil. While σ is likely to be similar within each soil, it is just as likely dissimilar between soil types. It has to be remembered that—statistically—the goal of analysis is an unbiased estimate of a specific parameter.

Parameters routinely measured in soil include: bulk density; particle-size distribution; soil-moisture characteristic curves; water flux; infiltration; water storage; hydraulic conductivity; and soil-water diffusivity. Here, we use bulk density as an example. If bulk-density samples are extracted from various depths within a research area, we might suspect—from a knowledge of statistical analysis and changes in soil texture with depth and space—that there is a distribution curve to describe the frequency of a given density to the number of observations or samples extracted. Figure 16.1 shows the frequency of distribution for bulk-density samples extracted from a clay soil. The normal frequency distribution is mathematically

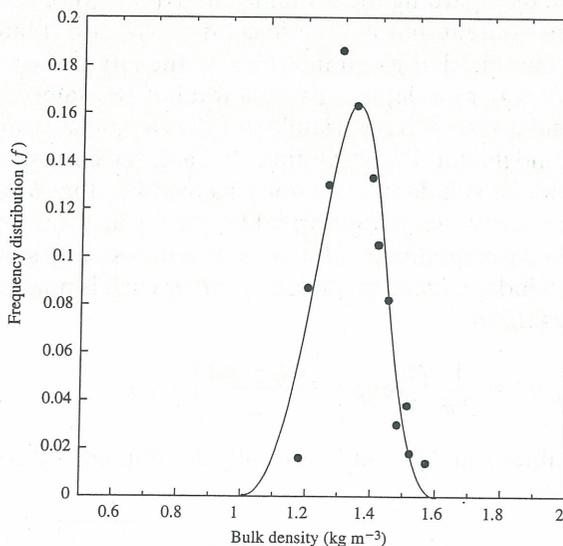


Figure 16.1 Measured (•••) and calculated (—) frequency distributions for soil-bulk density (data from Nielsen et al. 1973)

described by

$$f = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i - m)^2}{2\sigma^2}\right] \quad (16.3)$$

where f is the frequency. For a normal curve, the center position of the curve is determined by the mean of all samples extracted. Also, the less σ varies, the sharper and narrower the curve will be. Assuming a normal distribution, if a sample X were extracted at random from the population, the chance (probability) that it would fall between two other points x_1 and x_2 is expressed by

$$P\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} f dx \quad (16.4)$$

Typically, some soil properties are also log-normally distributed. The frequency function of a log-normal distribution is

$$f = \frac{1}{\sigma(x_i - \beta)\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x_i - \beta) - m]^2}{2\sigma^2}\right\} \quad (16.5)$$

for $x_i > \beta$ and $f = 0$ for $x_i < \beta$, where β is any constant to allow $\ln(x_i - \beta)$ to be normally distributed.

Several statistical methods are used to represent the variability of a given parameter on a relative basis such as the coefficient of variation $\{cv = (\sigma/m) * 100\}$. For example, both saturated-water content (θ_s) and bulk density have a low coefficient of variation when compared to textural analysis, and θ at varied soil pressures that have a medium coefficient of variation. Parameters such as saturated- and unsaturated hydraulic conductivity; pore-water velocity; apparent diffusion; and electrical conductivity all have a high coefficient of variation. As variability increases, more samples are needed for adequate analysis. Also, those parameters with high variability are likely log-normally distributed (owing to soil heterogeneity) while those of low variability have a normal distribution. As a result, most of the spatial variability within soil is described statistically. To prove this, we use hydraulic conductivity as an example, and because K is generally more variable than ρ_b .

Hydraulic conductivity is represented as a function of percentage of saturation, where the coefficient of variation generally increases with decreasing percentage saturation. Hydraulic conductivity is treated by regarding the sampling area as a homogeneous soil mass—in which case, both soil-water content and hydraulic conductivity are treated as two separate experimental variables, measured at given intervals. In the latter case, we suppose that 30 samples are taken from six separate depths, giving a total of 180 samples from the area. Using this approach, we consider the spatial variability of the sample area and wish to know if—for an average value of θ and its standard deviation—what is the corresponding hydraulic conductivity (mean value) and its standard deviation? Remember that θ most likely has a normal frequency distribution, and K has a log-normal frequency distribution. We can determine for any particular θ , the corresponding value of K (both at steady state) as well as its variability, by considering two independent variables, x_i and y_i ; each is normal and their joint distribution function is expressed as

$$F(x_i, y_i) = \frac{1}{2\pi} \iint \exp\left[-\frac{(x_i + y_i)^2}{2}\right] dx_1 dy_1 \quad (16.6)$$

By introducing two new variables x and y —both normally distributed and correlated— x and y can be related to x_1 and y_1 by

$$x = m_1 + \sigma_1 x_i \quad \text{also} \quad y = m_2 + r\sigma_2 x_i + \sqrt{1 - r^2} \sigma_2 y_i \quad (16.7)$$

These two equations are arranged such that the new variables x and y have means m_1 and m_2 , standard deviations σ_1 and σ_2 , and the correlation coefficient r . Rearranging, we have

$$x_i = \frac{(x - m_1)}{\sigma_1} \quad (16.8)$$

and

$$y_i = \frac{1}{\sqrt{1 - r^2}} \left[-r \left(\frac{x - m_1}{\sigma_1} \right) + \frac{y - m_2}{\sigma_2} \right] \quad (16.9)$$

Thus, the joint distribution function, as a function of the new variables, becomes

$$f(x, y) = \frac{1}{2\pi} \iint D \exp \left[-\frac{Q(x, y)}{2} \right] dx dy \quad (16.10)$$

where

$$Q(x, y) = \frac{1}{1 - r^2} \left[\frac{(x - m_1)^2}{\sigma_1^2} - \frac{2r(x - m_1)(y - m_2)}{\sigma_1\sigma_2} + \frac{(y - m_2)^2}{\sigma_2^2} \right] \quad (16.11)$$

and D is the Jacobian

$$D = \begin{vmatrix} \frac{\partial x_1}{\partial x} & \frac{\partial x_1}{\partial y} \\ \frac{\partial y_1}{\partial x} & \frac{\partial y_1}{\partial y} \end{vmatrix} \quad (16.12)$$

The joint frequency function of x and y (Cramer 1955) is

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}} \exp \left[-\frac{Q(x, y)}{2} \right] \quad (16.13)$$

This is the general form of the two-dimensional, normal frequency distribution function. By letting $x = \ln z$, we can rewrite the equation as

$$f(z, y) = \frac{1}{2\pi\sigma_1\sigma_2 z \sqrt{1 - r^2}} \exp \left[-\frac{Q(x, y)}{2} \right] \quad (16.14)$$

where z represents depth of sample and

$$Q(z, y) = \frac{1}{1 - r^2} \left[\frac{(\ln z - m_1)^2}{\sigma_1^2} - \frac{2r(\ln z - m_1)(y - m_2)}{\sigma_1\sigma_2} + \frac{(y - m_2)^2}{\sigma_2^2} \right] \quad (16.15)$$

Earlier we mentioned that K is log-normally distributed; by rewriting the log-normal distribution equation in terms of K at steady state (K_{ss}) and setting $\beta = 0$ in equation 16.5, we have

$$f_1(K_{ss}) = \frac{1}{\sigma_1 K_{ss} \sqrt{2\pi}} \exp \left[-\frac{(\ln K_{ss} - m_1)^2}{2\sigma_1^2} \right] \quad (16.16)$$

Also, because θ at steady state (θ_{ss}) is normally distributed, we can rewrite the normal frequency distribution (equation 16.3) as

$$f_2(\theta_{ss}) = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left[-\frac{(\theta_{ss} - m_2)^2}{2\sigma_2^2} \right] \quad (16.17)$$

The joint distribution function after equation 16.13 is

$$f(K_{ss}, \theta_{ss}) = \frac{1}{2\pi\sigma_1\sigma_2 K_{ss} \sqrt{1 - r^2}} \exp \left[-\frac{Q(K_{ss}, \theta_{ss})}{2} \right] \quad (16.18)$$

where

$$Q(K_{ss}, \theta_{ss}) = \frac{1}{1-r^2} \left[\frac{(\ln K_{ss} - m_1)^2}{\sigma_1^2} - \frac{2r(\ln K_{ss} - m_1)(\theta_{ss} - m_2)}{\sigma_1\sigma_2} - \frac{(\theta_{ss} - m_2)^2}{\sigma_2^2} \right] \quad (16.19)$$

Combining equations 16.17–16.19, the conditional frequency function of K_{ss} relative to another steady-state soil-water content (θ_{ss1}) is described statistically by

$$\frac{f_2(K_{ss}, \theta_{ss1})}{f_2(\theta_{ss1})} = \frac{1}{\sigma_1 K_{ss} \sqrt{2\pi} \sqrt{1-r^2}} \exp \left[-\frac{\left(\ln K_{ss} - m_1 - \frac{r\sigma_1}{\sigma_2} (\theta_{ss1} - m_2) \right)^2}{2\sigma_{1ss1}(1-r_{ss1})} \right] \quad (16.20)$$

Equation 16.20 is a normal frequency distribution function for K_{ss} with the mean value $m_3 = m_1 + r\sigma_1(\theta_{ss1} - m_2)\sigma_2^{-1}$ and the standard deviation $\sigma_3 = \sigma_1(1-r^2)^{1/2}$. By letting ε = arithmetic mean of K_{ss} and τ = corresponding standard deviation, the following relations are valid such that $\varepsilon = \exp[m_3 + \sigma_3^2/2]$ and $\tau = \exp[m_3 + \sigma_3^2/2] [\exp(\sigma_3^2 - 1)]^{1/2}$. As a result, we have an estimation of the mean and standard deviation of K_{ss} relative to any given soil-water content, based on frequency distributions. This statistical approach is an appropriate method for a field study. Consequently, we see that the spatial variability of soils and the samples extracted from them can be statistically treated, so long as an adequate number of samples is taken. The number of samples are determined, assuming a normal distribution, from the following equation

$$n = \frac{Z_\alpha^2 v_s}{d^2 m^2} \quad (16.21)$$

where n is the required sample size, Z_α is the value of the standardized normal variate corresponding to the level of significance α (the value of Z_α can be obtained from a cumulative normal frequency distribution table), v_s is the sampling variance, and d is the margin of error expressed as a fraction of the plot mean. The information of primary interest to the scientist is usually the treatment mean (the average over all plots receiving the same treatment), rather than an average of a single plot or treatment area. Hence, the desired degree of precision is usually specified in terms of the margin of error of the treatment mean, rather than of the plot mean. In this particular case, sample size is computed by

$$n = \frac{Z_\alpha^2 v_s}{r(D^2 m^2) - (Z_\alpha^2 v_p)} \quad (16.22)$$

where n is the required sample size, Z_α and v_s are as defined previously, v_p is the variance between plots of the same treatment (i.e., experimental error), and D is the prescribed margin of error expressed as a fraction of the treatment mean. As an example, suppose a researcher wishes to measure the hydraulic conductivity within a research plot using extracted, intact soil cores with a larger coring sampler. The researcher wishes to determine the number of samples necessary to achieve an estimate of the treatment mean within 5% of the true value. This researcher knows from previous studies that the following values can be used: $Z_\alpha = 1.95$, $v_s = 5.043$ (i.e., a *cv* of 28.4 percent), $D = 0.05$, $m = 25$ (the average number of samples the researcher has taken in the past with a smaller sampler), and $v_p = 0.1832$ —estimated from previous research. A sample size—number of cores required per plot—that can satisfy the researchers requirement at the 5% level of significance is computed as:

$$n = \frac{(1.95)^2(5.043)}{4(0.05)^2(25)^2 - (1.95)^2(0.1832)} = 3.45$$

Or, about four cores per research plot. Thus, by extracting an adequate number of samples, a great deal of spatial variability is nullified, and while calculated frequency distributions do not exactly fit measured values, they are a good approximation to work with when dealing with the effects of spatial variability on collected data. It is beyond the scope of this text to give a detailed treatise on statistics; the reader is therefore urged to consult a standard college text on statistics for a more in-depth discussion of the subject.

QUESTION 16.1

You have been given the following ten properties for an in-situ media: bulk density; water content at saturation; saturated and unsaturated conductivity; pore-water velocity; diffusion coefficient; particle-size analysis; water content at -100 kPa; and the scaling coefficient. For which properties would you expect a normal distribution?

Geostatistics

The frequency distribution just discussed typically deals with statistics as associated with univariate data and hypothesis testing. In some cases, it can extend to a correlation between two variables and linear regression. Generally, in the discussion of correlation between variables it is usually two different attributes that are considered. However, in the earth sciences we often wish to make a treatment of the correlation between values of a single variable, measured at different points in space. The ability to analyze many such measurements against a spatial framework is a necessity, and is where standard statistics are inefficient for obtaining solutions to more complex problems.

For example, consider typical histograms of two data sets. Visually, we see little difference between the two, and a Kolmogorov–Smirnov test for comparisons—from which the sample data comes—would not reject (with a 5 percent significance level) a null hypothesis of no difference between the two populations. However, suppose we have contour maps of the same data. The first set of data shows a normal contour map, while the second set of data produces a contour map that is “busier”—so how is it possible that the histograms look the same? They look the same because the data were collected at the nodes of a regular two-dimensional grid, but it is within the grids where the contour map shows one data set busier than the other. Thus, the difference between the two contour maps is a reflection of the more random spatial arrangement of data in the busier map; in other words, there is less correlation between adjoining data pairs in the busier contour map. This is where geostatistics shows a difference and hence, is more valuable as an analysis tool than statistics alone.

One of the first scientists to recognize the necessity of accounting for spatial correlation between data was D. G. Krige (Journel and Huijbregts 1978). Krige derived empirically based “regression weights” that could be applied to the grades of channel-ore samples used in the estimation of slopes. Based on this work, others undertook a formal development of his theories, and the field of geostatistics was born. As a simple definition, geostatistics is the statistics of spatial- (or temporally) correlated data. It enables the scientist to measure spatial autocorrelation and evaluate the nature and quality of raw data. Included in this discussion is the term spatial variability; this is a common term used to indicate that geologic media and soils—most of which are heterogeneous—change with space and time. Such properties include bulk density, porosity, soil texture, water content, pH, and hydraulic conductivity. Because soils are spatially variable, it is important to have tools with which to measure that variability. We have already discussed some of these tools, but there are others that we now describe in our continuing discussion of geostatistics. (For an excellent treatise on spatial variability, the reader is referred to chapter 13 of Hillel 1980.)

Semivariogram

Geostatistics incorporates data taken from a grid (usually square), with a unit spacing. Upon analyzing the data, a histogram and contour map are usually generated. To describe the spatial correlation between samples in near proximity, a semivariogram is used. This is a basic geostatistical tool that allows us to visualize, model, and exploit the spatial autocorrelation of a regionalized variable. The function of the semivariogram is half the average squared difference between data pairs of points that are separated by displacement, \vec{h} . This can be calculated by

$$\gamma^*(\vec{h}) = \frac{1}{2N(\vec{h})} \sum_{i=1}^{N(\vec{h})} [Y(\vec{x}_i + \vec{h}) - Y(\vec{x}_i)]^2 \quad (16.23)$$

where N is the number of data pairs in the region, and Y represents the value of the data at location \vec{x}_i .

A semivariogram needs to be calculated for a variety of directions, to allow recognition of anisotropic variability; for example, east to west and north to south. Commonly, neighboring sample pairs are closer in value than more separated pairs. Once an experimental semivariogram is obtained, it is modeled to obtain block estimates; usually, a spherical model is used. However, there is no one “correct” model for a particular situation, so the semivariogram of choice determines the amount of smoothing necessary in later steps. The accuracy of the modeling process depends on the number of data pairs used in the calculations; on the experimental semivariogram; and on the lag distance at which it is evaluated. The modeling process is complicated and can involve: polygonal estimation; inverse-distance weighting; inverse-distance-squared weighting; estimation variance; confidence limits; as well as other factors. Since this chapter is an introduction to various tools used in the earth sciences, discussing these parameters is beyond its scope; the reader is referred to Hohn (1988), Wackernagel (1995), or other texts on geostatistics.

Kriging

We now understand that it is possible to rank any estimation regime to its efficiency by calculating the estimation of variance. However, for this we need to determine the best set of weights for a particular block-sample configuration—that is, the set of weighting coefficients that minimize the estimation of variance. The process of calculating this optimal estimation of variance is called “kriging,” named after D. G. Krige, who pioneered the work of geostatistics. Kriged estimates are used for drawing a contour map, such as for water-table depth. Since geostatistics are generally done on computers, the basics steps to computer contour mapping using the kriging technique are: **(1)** collect data to estimate the variance; **(2)** superimpose data on a regular grid; **(3)** interpolate values at each node on the grid; **(4)** construct contours; **(5)** smooth the contour lines by splining or other techniques, if necessary; and **(6)** draw the contour map.

Beginning, we assume the regionalized variable under study has the value $z_i = z(x_i)$, each representing the value at the point x_i . We also assume that this regionalized variable is a second-order stationary, with expectation E

$$E\{Z(x)\} = m \quad (16.24)$$

where $Z(x)$ is the random variable observed at point x and m is the first-order moment. Three second-order moments are also useful in geostatistics; these include the variance of the random variable, the covariance, and the semivariogram function (discussed earlier). This is generally estimated by a centered covariance C

$$E\{Z(x+h)Z(x)\} - m^2 = C(h) \quad (16.25)$$

and a variogram

$$E\{[Z(x+h) - Z(x)]^2\} = 2\gamma(h) \quad (16.26)$$

where h is the distance from point x . The Kriged estimator—that is, a linear combination of n values of the regionalized variable—is

$$z_{k^*} = \sum_{i=1}^n \lambda_i Z_i \quad (16.27)$$

where λ is the calculated weight. This ensures that the estimate is unbiased, and the estimation variance minimized. When the kriging theory is met, we quickly interpolate values of the sampling variable between measured points; then contour lines are drawn. Kriging allows a determination of the estimation of variance, useful for determining uncertain values of the function, as well as in subsequent sampling of the same area within the field. However, the calculation of λ and other parameters are beyond the scope of this text, so the reader is once again referred to Hohn (1988) and Wackernagel (1995) for an in-depth treatise on Kriging and geostatistics.

Because even the simplest calculations in geostatistics are performed on the computer (they are tedious and unwieldy for large data sets otherwise), the practitioner needs to be well-versed in some computer programs, as well as a variety of techniques in problem solving; these include: univariate statistics; multivariate statistics; means; histograms; scattergrams; semivariograms; variograms; interactive curve fitting; plotting; grid searching; equation solving; contouring; and map drawing. We have mentioned only the main parameters for a basic understanding of geostatistics here; others might include variogram cloud; variogram and covariance function; extension and dispersion variance; measures and plots of dispersion; linear model of regionalization; Kriging spatial components; and many more.

A concluding note on geostatistics: while it is widely used in the earth sciences, it is not the only tool for solving complex problems with large data sets. However, it is easier to understand intuitively than other tools that we now discuss: scaling and fractals.

Power-Law Distributions

Random variables The concept of random variables is basic to modern statistics. A function of a variable is a rule (mathematically speaking) whereby one or more numerical values are associated with different values of a variable. For example, consider the function $f(x) = 2x + 3$. If $x = 0$, then $f(x) = 3$, if $x = 2$ then $f(x) = 7$, and so forth. Assuming x is a random variate, the probability density function of x , $f(x)$, gives the probability that the variate assumes the value x , that is, $f(x) = Pr(x)$. If we distinguish between the name of the variate (x) and the values the variate assumes, then $f(x = x') = Pr(x = x')$. Usually, $f(x)$ represents a model for the relative frequency (in the series of experiments), with which the variate x assumes specified values. These values normally vary both spatially and temporally, and each quantity measured is termed a random variable. Since these values fluctuate, we need to associate the random variable with a probability distribution, $F(x)$. The probability that the random variable x assumes a specific value x_j is given by $Pr(x = x_j) = f(x_j)$. The distribution function, as a probability law, is given as

$$Pr(x \leq x_j) = F(x_j) = \int_{-\infty}^{x_j} f(x) dx \quad (16.28)$$

Thus, the distribution function of a random variable x represents a cumulative probability. Any nonnegative function, whose integral over the entire range of the variate in the function is unity (1), defines a probability density. Consequently, a random variable x is said to have a

density function $f(x)$ if

$$\int_{-\infty}^{x_j} f(x) dx = Pr(x \leq x_j) \tag{16.29}$$

In science, the variation we measure is often defined as “measurement error.” It then becomes important to overcome variability due to sampling and environmental problems. As a result, we need to use distribution functions to help interpret the data gathered. There are a number of distribution functions for random variables that we discuss in reference to power-law distribution, but for exact, technical detail on the evaluation of data fit to a particular distribution function, we refer the student to a standard statistics text. An example of a cumulative probability distribution is shown in figure 16.2.

For any given random variable, there is a probability distribution associated with it. If we let the symbol x denote such a random variable, the symbol $f(x)$ denotes the probability density for x . Suppose the probability of the random variable x assumes the specific x_i , this can be expressed as

$$Pr(x = x_i) = f(x_i) \tag{16.30}$$

It is important to remember that a probability measure on the random variable x is defined by a function that has the following properties:

$$0 \leq Pr(x \leq x_i) \leq 1$$

$$Pr(-\infty \leq x \leq \infty) = 1 \tag{16.31}$$

$$\text{for } x_j > x_i \quad Pr(x_i \leq x \leq x_j) = Pr(x \leq x_j) - Pr(x \leq x_i)$$

The distribution for the random variable x is denoted by $F(x)$. As a probability law, the distribution is interpreted as that given in equation 16.28. In this instance, $f(x) dx$ is the product and defines the area of a rectangle with height $f(x)$ and width dx , and is called the probability element. It is normally convenient to assign the range of a random variable as $\pm\infty$. Thus, the density function is zero for all values of $x \leq 0$ and the probability density is defined

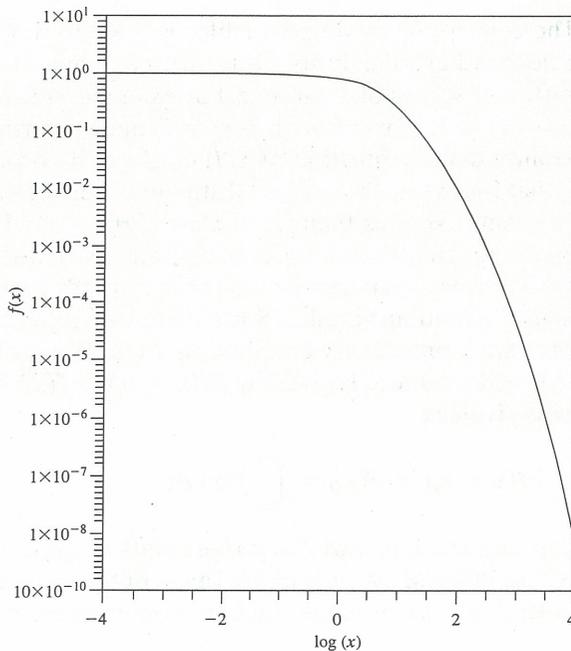


Figure 16.2 A cumulative probability distribution

as $Pr(x) = f(x)$, hence $Pr(x) = 0$. The random variable is either continuous or discrete. The distribution of a random variable is generally associated with parameters that are constants, determining certain characteristics of the distribution. An example of the density function and its relation to the distribution function is given in figure 16.3a. In figure 16.3b, the density at the point $x = x_j$ is given by the height $f(x_j)$.

Log-normal distribution For log-normal distributions, the variable is a result of multiplication rather than a sum, as is the case for normal distributions (e.g., bulk density). Many natural phenomena tend to follow log-normal distributions since the log-normal distribution is not symmetric, and is defined only for positive x . For example (using hydraulic conductivity), suppose a researcher is extracting soil cores to obtain an “average” value of K over a specific area. Variability in each sample can include soil type, bulk density, occluded pore space, presence of cracks, or particle size. To obtain an average hydraulic conductivity, the value of K for each core is summed for the whole. However, each of the factors of variability influences the individual K value. Each value is therefore subjected to variability based on the physical factors listed as well as the K value obtained for each core, and for all cores is proportional to the product of all factors involved. As a result, the process is multiplicative and the value obtained for K for each core is likely distributed log-normally.

For a probability density function (pdf), the log-normal pdf can be expressed as

$$P(x) = \frac{\exp\left[-\frac{(\ln x - \alpha)^2}{2\beta^2}\right]}{x\beta\sqrt{2\pi}} \quad (16.32)$$

Because x can be expressed only positively, $P(x)$ approaches zero as x approaches zero. Since the expected value of x , $E[x] = \exp(\mu + \sigma^2/2)$, and the sampling variance, $\text{Var}[x] = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$ then, the coefficient of variation, CV is given by the expression

$$CV = [\exp(\sigma^2) - 1]^{0.5} \quad (16.33)$$

We write the cumulative log-normal distribution as

$$P[x] = \int_0^x e^{-[\ln(x) - \mu]^2/2\beta^2} \frac{dx}{\sqrt{2\pi}\beta x} \quad (16.34)$$

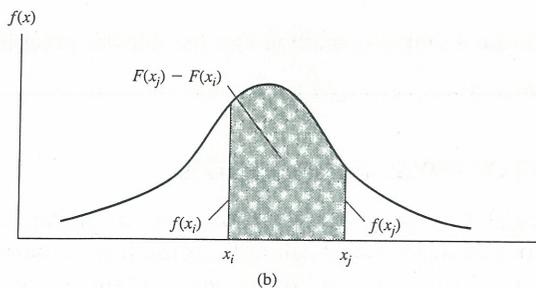
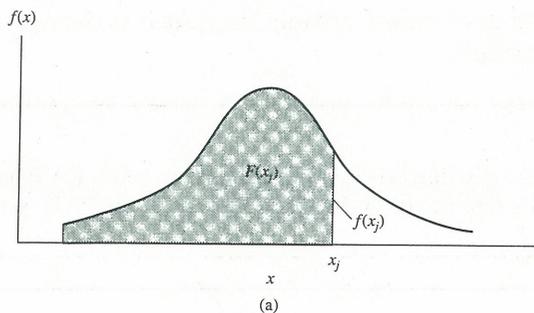


Figure 16.3 (a) An example of the density function, $f(x)$, and its relation to the distribution function, $F(x)$. (b) The area under the curve corresponds to the probability, $F(x_j) - F(x_i)$.

For this case, the normal curve of error is mathematically expressed by

$$P[x] = \frac{1}{\beta} N\left(\frac{\ln(x) - \mu}{\beta}\right) \quad (16.35)$$

QUESTION 16.2

For the properties given in question 16.1, how would you represent the remaining soil properties? What makes a log-normal distribution more appropriate than a normal distribution for certain physical properties?

Exponential distribution If $f(x) = \gamma e^{-\gamma x} = 0$ (where both x and γ are greater than zero), the random variable is x , and the parameter is γ . We make the explicit distinction between the variable and the parameter by writing $f(x|\gamma) = \gamma e^{-\gamma x}$. For $\gamma = 1$, the probability density is written as $f(x|\gamma = 1) = e^{-x}$ (where $x > 0$). Consequently, the random variable defined by this density is called an “exponential variable.” Also, the exponential distribution is really a *family* of distributions; an individual member can be specified by assigning a numerical value to the parameter γ . The distribution function for an exponential variable is mathematically expressed as

$$Pr(x \leq x_i) = \int_0^{x_i} f(x) dx = \int_0^{x_i} \gamma e^{-\gamma x} dx \quad (16.36)$$

An example of an exponential distribution is the Poisson distribution. Suppose we want to study the emission of fast-moving neutrons from a neutron probe, for measuring soil volumetric-water content. We can investigate either a single event or no event, with related probabilities of $\lambda \Delta x \ll 1$ and $1 - \lambda \Delta x$. In this study, Δx is so short that it does not contain more than one event. Thus, if we consider a set interval Δx , the number of events within this interval $N(\Delta x)$ will follow the Poisson distribution, which can be expressed mathematically as

$$Pr[N(\Delta x) = k] = \frac{(\lambda \Delta x)^k e^{-\lambda \Delta x}}{k!} \quad (16.37)$$

where $k = 1, 2, 3, \dots$, $Pr[N(\Delta x) = k]$ is the probability of finding k events within Δx (also written as δx), and λ is the number of events per unit time, such as the neutrons being emitted from a probe (assumed constant).

QUESTION 16.3

(a) Explain the difference between sampling error and measurement error. (b) What is the difference between statistical true value and scientific true value? (c) between scientific bias, measurement bias, and sampling bias?

QUESTION 16.4

Why would someone purposely choose a sampling method that has a lower precision over a method having a higher precision?

16.2 SCALING AS A TOOL FOR DATA ANALYSIS OF PHYSICAL PROPERTIES

Scaling is a physical or geometrical difference between soil types or parameters. It is also defined as a statistical difference between soils and related parameters; a mathematical definition can also be applied. Miller and Miller (1956) introduced the “similar media” concept. In

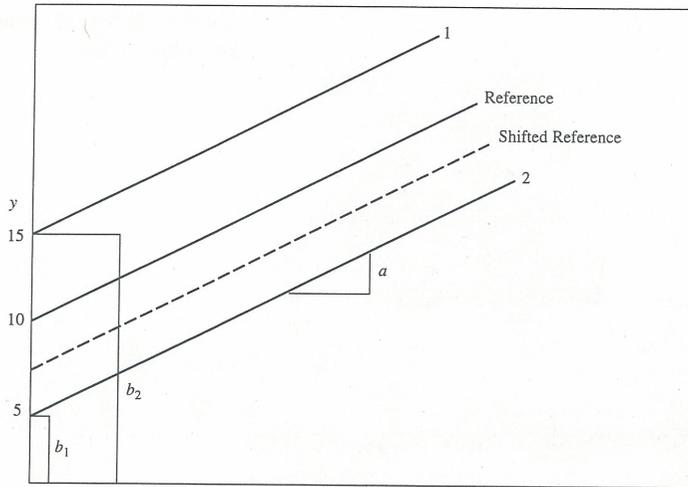


Figure 16.4 Example of shifted reference curve

principle, this concept allows description of soil-water behavior in one soil from either experimental or computed data; in another by employing reduced variables defined in terms of appropriate microscopic characteristic lengths. Basically, similar media differ only in the scale of their internal microscopic geometries, and thus have equal porosities—that is, scaling.

Scaling is used to simplify the description of the statistical variation of soil properties. By this simplification, the distribution of spatial variation is described by a set of scale factors α_r , relating the soil hydraulic properties at each location r to a representative mean.

Tillotson and Nielsen (1984) refer to functional normalization as an empirical method to determine scale factors. Its objective is to coalesce all relations in the set into a single reference curve that describes the set as a whole; the resulting scale factors have no physical significance. For example, consider two curves, 1 and 2 (see figure 16.4):

$$\begin{aligned} y_1 &= ax + b_1 \\ y_2 &= ax + b_2 \end{aligned} \quad (16.38)$$

where x and y are variables and a and b are constants. The reference curve is defined such that

$$y_{\text{ref}} = ax + b_{\text{ref}}; b_{\text{ref}} = y_{\text{ref}} - ax \quad (16.39)$$

and

$$\alpha_1 = \frac{b_{\text{ref}}}{b_1}; \quad \alpha_2 = \frac{b_{\text{ref}}}{b_2} \quad (16.40)$$

then

$$\begin{aligned} y_1 &= ax + \frac{b_{\text{ref}}}{\alpha_1} \quad \therefore \quad y_1 = ax + y_{\text{ref}} - ax \\ y_2 &= ax + \frac{b_{\text{ref}}}{\alpha_2} \quad \therefore \quad y_2 = y_{\text{ref}} \end{aligned} \quad (16.41)$$

By performing this operation, both curves are coalesced to the reference curve (see figure 16.4), or a distribution set μ_α . For example:

$$\left. \begin{aligned} b_1 &= 15 \\ b_2 &= 5 \\ b_{\text{ref}} &= 10 \end{aligned} \right\} \left. \begin{aligned} \alpha_1 &= 10/15 = 2/3 \\ \alpha_2 &= 10/5 = 2 \end{aligned} \right\} \mu_\alpha = 4/3 \quad (16.42)$$

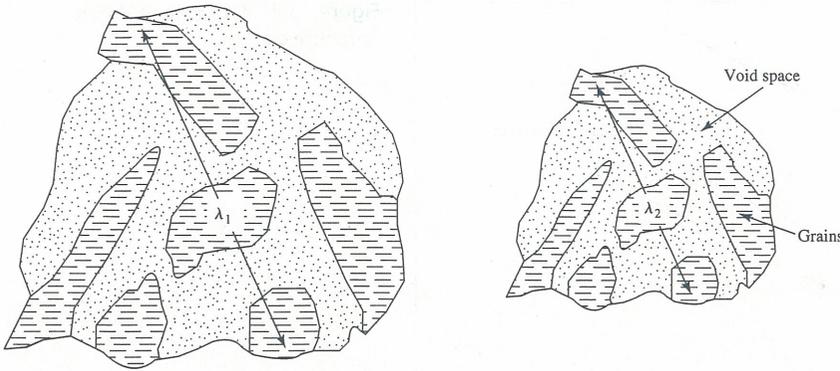


Figure 16.5 Similar media (after Miller and Miller 1956)

Now, by normalizing the α -distribution set $\mu_\alpha = 1.0$ by

$$\alpha_i = \frac{N\alpha_i}{\sum_{i=1} \alpha_i} = 8/3 \tag{16.43}$$

thus, evaluating equation 16.27 at a distribution set, $\mu_\alpha = 1.0$ and b_{ref} ,

$$\left. \begin{aligned} \alpha_1 &= \frac{2 \cdot 2/3}{8/3} = \frac{1}{2} \\ \alpha_2 &= \frac{2 \cdot 2}{8/3} = \frac{3}{2} \end{aligned} \right\} \begin{aligned} \mu_\alpha &= 1.0 \\ b_{ref} &= \alpha_1 b_1 = 7.5 \end{aligned} \tag{16.44}$$

In doing so, we have shifted the reference curve (see figure 16.4)!

Similitude analysis (similar media) gives physical significance to the scale factors. Miller and Miller (1956) introduced microscopic length, and according to their concept, soils are similar if they are geometric scales of each other—that is, all the microscopic geometric details of the one medium could be multiplied by a constant to obtain the microscopic details of the other medium. The microscopic length characteristic is denoted by λ and could be described as: (1) average grain size; (2) average pore diameter; (3) maximum grain size; (4) maximum pore size; and (5) also, by combinations of 1 + 2 and 3 + 4. An example of similar media is depicted in figure 16.5. For further examples, we look at both capillary rise (described by the Laplace equation) and Poiseuille’s law.

From the Laplace equation

$$\Delta P = \rho gh = \frac{2\sigma}{r} \tag{16.45}$$

where P is pressure, ρ is fluid density ($\text{m}^3 \text{kg}^{-1}$), g is gravitational force constant, h is height of rise (L), σ is surface tension (mN m^{-1}), and r is the maximum radius (L) of the water-filled pore. Equating r with λ , the scale factor between the two soils is given by $\alpha = \lambda_2/\lambda_1$. If $\alpha = 2$ —the maximum water-filled pore radius of medium 2 is $2x$ as large as that of medium 1—then ΔP is $2x$ as small, or $|h|$ is $2x$ as small. Corresponding retention curves for similar media 1 and 2 using this concept gives equation 16.46, with a graphical representation shown in figure 16.6.

$$h_1 \lambda_1 = h_2 \lambda_2; \quad h_r = \frac{h_m}{\alpha_r} \tag{16.46}$$

where h_r is reference pressure head, h_m is the average pressure head, and α_r is the scaling coefficient. From Poiseuille’s law we have:

$$K_r(\theta) = r^2 * f(n, l, \Delta P, g) \tag{16.47}$$

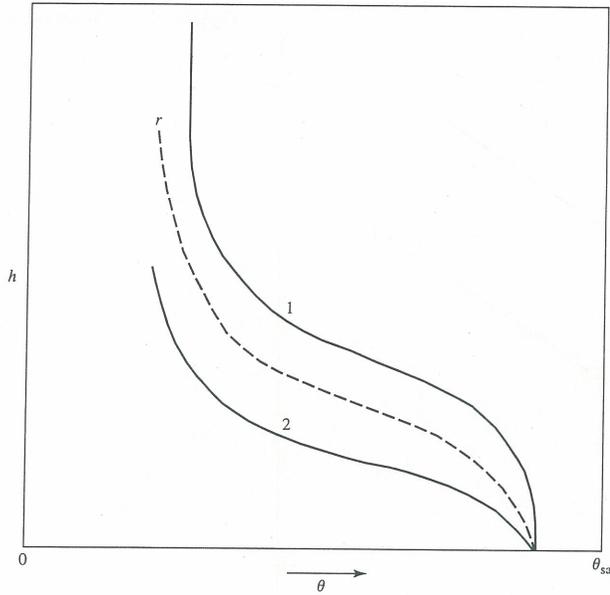


Figure 16.6 Corresponding moisture-characteristic curves for similar media 1 and 2

where K_r is the hydraulic conductivity at point r (i.e., the reference), n is the number of pore classes through which flow takes place, and l is the length of the material through which the fluid flows. In this case, if $\alpha = 2$, then K of medium 2 will be $4x$ as large as that of medium 1, and

$$\frac{K_1}{\lambda_1^2} = \frac{K_2}{\lambda_2^2}, \quad K_2 = \left(\frac{\lambda_2}{\lambda_1}\right)^2 * K_1 \quad (16.48)$$

or

$$K_i = \alpha_i^2 K_r \quad (16.49)$$

where the scale factor α_i is the characteristic length of medium i and (see figure 16.7).

From equation 16.46 and 16.49, we derive for sorptivity (S) such that

$$\begin{aligned} S_i &= \sqrt{K_i h_r(\Delta\theta)} \\ S_r &= \sqrt{K_r h_r(\Delta\theta)} \end{aligned} \quad (16.50)$$

$$S_i = \sqrt{\alpha_i^2 K_r \left(\frac{1}{\alpha}\right) h_r(\Delta\theta)}; \quad S_i = S_r \alpha_i^{1/2}$$

In other words, soil-water transport characteristics of a set of soils are connected through scale factors. Thus, by collecting retention data of a soil set, we automatically infer conductivity, diffusivity, and sorptivity relations. An example of scaling of $\theta(h)$ and $K(\theta)$ is shown in figure 16.8 (Warrick, Mullen, and Nielsen 1977). This is physically described by setting $r = 1, \dots, R$ locations and $i = 1, \dots, I$ pressure increments for which $\theta(h)$ is determined, such that $s = \theta/\theta_{\text{sat}}$ where $0 \leq s \leq 1$ (similar media: equal porosity) and $h_m = \alpha_r h_r$; thus

$$\frac{\alpha_1 + \alpha_2 + \dots + \alpha_R}{R} = 1.0 \quad (16.51a)$$

$$h_r = \log h_m = a_0 + a_1 S + a_2 S^2 + a_3 S^3 \quad (16.51b)$$

$$SS = \sum_{r,i} [h_m(S_i) - \alpha_r h_r(S_i)]^2 \quad (16.51c)$$

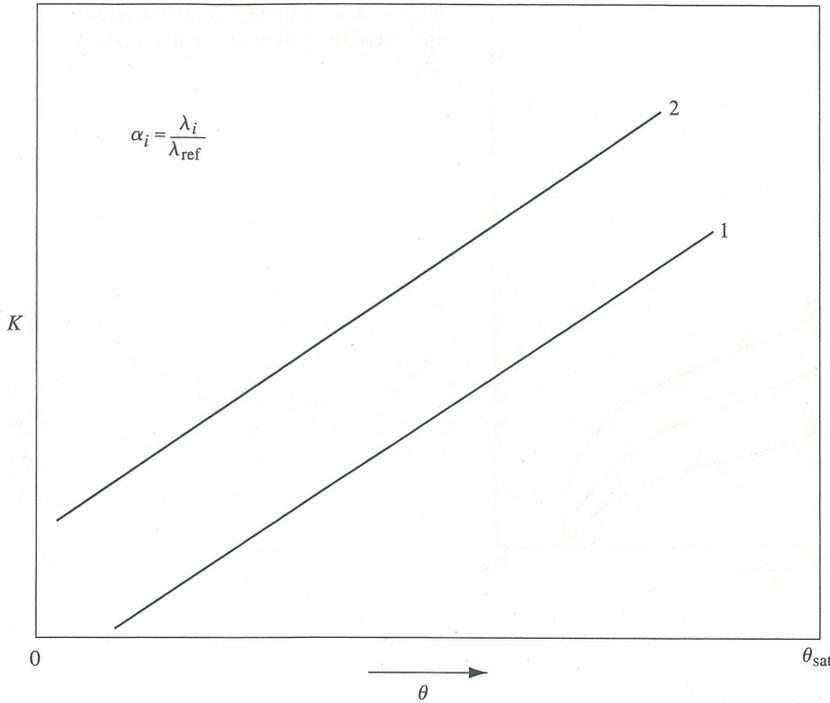


Figure 16.7 Scale factor α_i for K versus θ

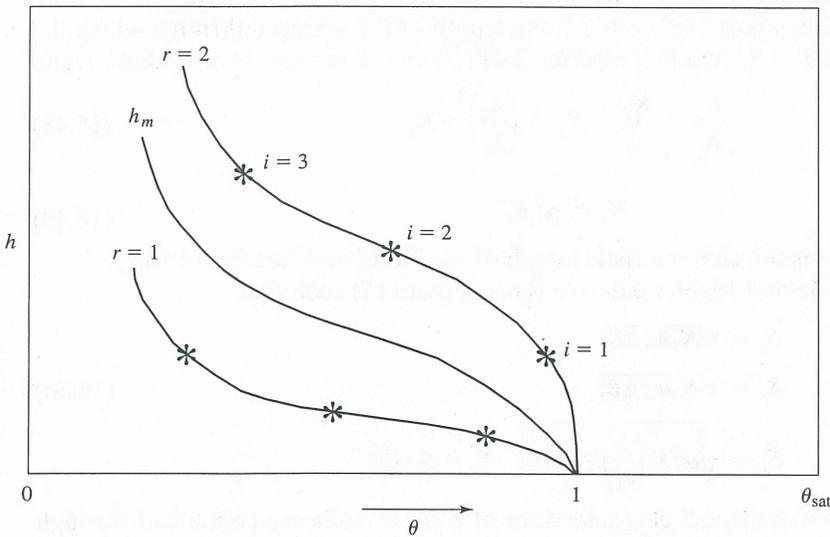


Figure 16.8 Example of scaling $\theta(h)$

where equation 16.51b is the reference curve. Further, we want to minimize SS as close to zero as possible; if it's zero, then SS is exact so scaling is exact. The basic steps are: **(1)** estimate $h_m(s)$ by multiple linear regression technique, that is, SAS, LMSL, and so on; **(2)** find α_r by minimizing equation 16.51c subject to 16.51a; **(3)** find new $h_m(s)$ such that

$$\log [h_m(S_i) = \log [\alpha_r h_r(S_i)] = a_0 + a_1 S(i) + a_2 S(i)^2 + a_3 S(i)^3 \quad (16.52)$$

and **(4)** repeat steps 2 and 3 until SS no longer changes.

Following this brief introduction to scaling, a generalized definition is helpful in understanding the concept. To that end, we use that of Shouse et al. (1989): “Scaling is a systematic

method for specifying a change of variable that transforms one system to another one with more desirable traits.” Scaling works very well for many problems where the soil is homogeneous on a large scale. Additionally, scaling can be applied to systems where there is a separation of scales. For example, a soil can be heterogeneous at a small scale, but if viewed at a large scale, it appears homogeneous. The reason for this is that the averaging is large compared to the scale of heterogeneity.

In addition to using scaling as a tool for transforming systems, we now develop new conceptual models by using the concepts provided by fractal mathematics, especially in the way we investigate heterogeneity. For example, Feder (1989) uses the coastline of Norway to illustrate length in the fractal dimension. Every time the resolution is increased on the coastline, there is a subsequent increase in its measured length—that is, the length continues to increase on at least several smaller “recursive” levels. Because the outline of the coast is heterogeneous, this model is referred to as a “self-similar” process. A primary characteristic of the coastline of Norway (or of fractals) is that the image we observe generally looks the same at almost any scale in the real world. This property is referred to as “fractal scaling.” Thus, scaling and fractals are closely related. The main advantage of fractal scaling is that traditional scaling techniques cannot be applied to materials that exhibit fractal scaling—such as physical description of topographic relief—to the development of consistent theories on both the formation and nature of fluid turbulence. Since a great number of shapes and processes directly affect soil and geologic formation, it is logical to apply the tools of fractal mathematics to the variability of hydraulic properties of soils and aquifers.

16.3 THE FRACTAL DIMENSION

This section is written to provide a brief introduction and overview of fractals, and some insight on how fractals facilitate the description of fluid flow and contaminant transport through soil, as well as other physical processes in soil. As a tool, it is much more suitable for these parameters than geostatistics. A complete description of fractals is not the objective of this text, so the reader is referred to the References section (at the back of the book) to gain more knowledge on the subject. Benoit Mandelbrot proposed the concept of fractal geometry in 1975 and since that time, fractal geometry has been used to render drawings of the surface of the planet Mars, trees, fractal shapes, paintings, movie scenes using computer graphics, and fluid flow and displacement. Indeed, computer programs using fractal generation have even been written to portray entire cities accurately. By running these programs, an individual can visit a given city via computer, and when he or she travels to that city in person, is already familiar with major streets, parks, restaurant locations, and other areas. The accuracy is phenomenal.

Prior to the development of fractal geometry, traditional Euclidean geometry was used to describe objects. What is the difference between the two forms of geometry? A comparison reveals that Euclidean geometry is based on a characteristic size or scale; works well in describing man-made objects; and is described by formulas—while fractal geometry is independent of scale; appropriately describes natural shapes; and is described by recursive algorithms. What is a fractal? By definition, “a fractal is a set for which the Hausdorff–Besicovitch dimension strictly exceeds the topological dimension” (Mandelbrot 1982). However, since this definition is a bit rigorous for the uninitiated, a more aptly put, simple definition is that fractals are repeating patterns—that is, by looking at the whole, many small parts, similar in appearance to each other and to the whole, make it up. Some of the best examples of fractals in nature are clouds, trees, and mountainous landscapes. With the advent of fractal geometry, during the last decade we have incorporated fractal mathematics with both computer and natural sciences, and it has quickly become a necessary tool in physics,

soil science, hydrology, and other natural sciences. Applications of fractals in soil science have recently been reviewed by Perfect and Kay (1995).

Triadic Von Koch Curves

The triadic von Koch curve (often called the snowflake curve) is a good example of the fractal dimension (D), where $D > 1$. The construction of this curve is either recursive or iterative. A line segment (termed the initiator) is divided into thirds, while the middle segment is replaced with two equal segments, forming an equilateral triangle. This process yields a line segment composed of $4/3$, called the generator. This generator segment is iteratively applied to itself to generate the von Koch curve (see figure 16.9). At the beginning, the initial straight-line segment has $n = 0$, while the generator has $n = 2$ (once applied); before application, $n = 1$ for the generator. Thus, at any iterative portion of the generated curve at any stage n , we have a prefractal; by applying a reduced generator to all segments of a generation of the curve, a new generation is obtained—such a curve is called a prefractal and, each small portion, when magnified, reproduces a larger portion, exactly. As a result, the curve is invariant under changes of scale, much like a coastline or outline of a rough rock. Generation of the curve on a computer can squeeze an infinite length into a finite area without the curve intersecting itself. This denotes the concept of self-similarity and is one of the fundamental, central properties of fractal geometry.

Self-Similarity and Scaling

Because the original line segment is invariant with respect to both translation and scaling, the expression for D is easily obtained. For the von Koch curve, the length of the prefractal (the n th generation) is expressed mathematically as

$$L(\delta) = \left(\frac{4}{3}\right)^n \quad (16.53)$$

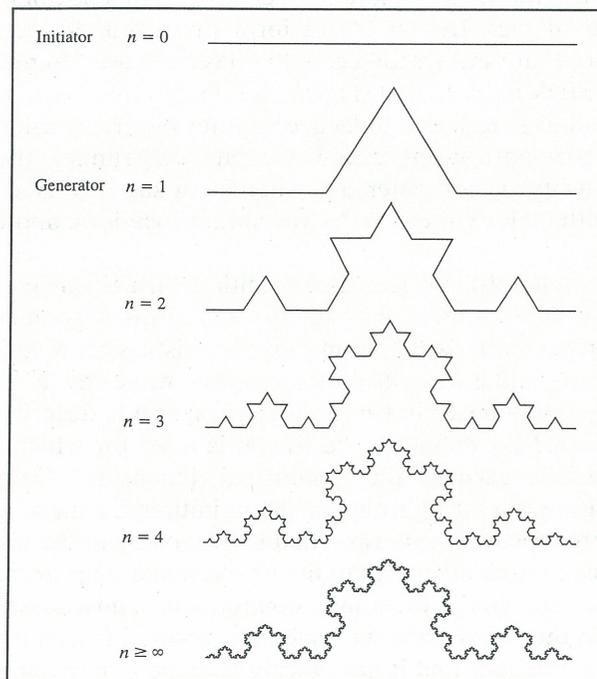


Figure 16.9 Example of the triadic von Koch curve showing the initiator and generator (data from Peitgen and Saupe 1988)

where L is the line segment of length (δ) and n is the iterative step of the curve generation. The length of each small segment of the line can be expressed as $\delta = 3^{-n}$, thus n (the generation number) is written as

$$n = -\frac{\ln \delta}{\ln 3} \quad (16.54)$$

Consequently, the length is expressed as

$$L(\delta) = \left(\frac{4}{3}\right)^n = \exp\left(-\frac{\ln \delta [\ln 4 - \ln 3]}{\ln 3}\right) = \delta^{1-D} \quad (16.55)$$

where $D = \ln 4 / \ln 3 \approx 1.26$. We state that the number of segments in the line is $N(\delta) = 4^n = 4^{-\ln \delta / \ln 3}$ and is written in the form $N(\delta) = \delta^{-D}$, where D is the fractal dimension, 1.2628 for the triadic von Koch curve. As we see, unlike the more familiar Euclidean dimension, the fractal dimension need not be an integer.

If points are specified in some space, using the Cartesian coordinate system, the location of a line drawn through a certain point can be determined. Likewise, if the length scale is changed in the positive direction by the same factor (λ) for all components of x , then a new system of points can be mapped. Also, if the second set of points is adjusted by a factor $(1 - r)x$, the original set of points can be retrieved. With this reasoning, a plane is invariant under translation in that plane, as well as uncertain change of scale or length of scale. There is also statistical self-similarity in which, upon magnification, the segments of a line look alike, but are never exactly alike at different scales. For example, if we consider a coastline, as did Feder(1989), the more closely we follow the smaller indentations or curves, the longer the coastline becomes. In this case, each smaller section of coastline has the same appearance of the whole coastline, but not exactly. Thus, the total length of the coastline is the yardstick δ multiplied the number of measurements of size λ , $N(\delta)$, in measuring the coastline—that is, coastline length = $\lambda \cdot N(\delta)$, where $N(\delta)$ varies on the average of δ^{-D} and length: $\alpha \lambda \cdot \delta^{-D} = \delta^{D-1}$. Here, $D > 1$ and as the length of the yardstick used to measure the coastline length decreases, coastline length increases. For real coastlines D is about 1.15 to 1.25. Consequently, the similarity dimension D_s is

$$D_s = -\frac{\ln N}{\ln \lambda(N)} \quad (16.56)$$

This similarity dimension is relatively easy to determine for self-similar fractals such as the von Koch curve and its variants. The basic values to know for fractal dimensions are those for the set of points that make up a line in ordinary Euclidean space $D = 1$, for the set of points that form a surface in space $D = 2$, and for a ball or sphere $D = 3$.

Considering the von Koch curve to be the graph of a function $f(t)$, a scaling ratio of $\lambda = (1/3)^n$, where $n = 0, 1, 2, \dots$, the property of the Koch curve is $f(\lambda) = \lambda^\alpha f(t)$ where the scaling component $\alpha = 1$. Because $f(t)$ is not a single value, the scaling relation is true for any point within the set. The power-law function $f(t) = bt^\alpha$, the same type of construction used on functions defined over all real positive numbers, satisfies the homogeneity relation $f(\lambda t) = \lambda^\alpha f(t)$ for all positive values of the scale factor λ . This function, and functions that satisfy this relation, are termed scaling.

Box Dimension

The fractal dimension of the coastline example is determined by covering it with a set of squares of edge length δ —the unit of length equal to the edge of the box width. The number of squares required to cover the coastline yields $N(\delta)$, from which the fractal dimension is

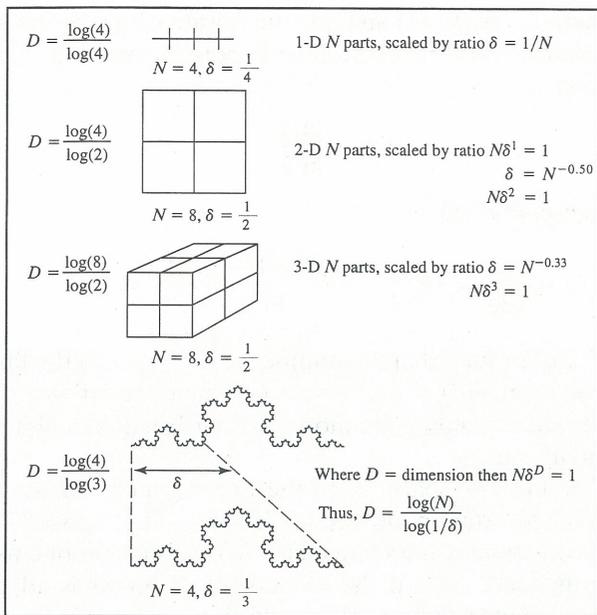


Figure 16.10 The box dimension and its various components (data from Peitgen and Saupe 1988)

determined by finding the slope of $\ln N(\delta)$, plotted as a function of $\ln \delta$ (see figure 16.10). This is an alternative definition of the fractal dimension, and is useful because the similarity dimension is of no value for either statistically self-similar or scale-invariant objects; this is because the “box” dimension measures how much space is filled by a geometrical object. The number of boxes required to cover a set S , if the entire S is contained within one box of size δ_{\max} , is given by

$$N_{\text{box}}(\delta) = \left(\frac{\delta_{\max}}{\delta} \right)^D \tag{16.57}$$

or $N_{\text{box}}(\delta)$ is proportional to δ^{-D} (figure 16.10). This definition of box dimension is one of the most useful methods for estimating fractal dimension. For small sets, δ can be greater than the spacing between specific points, but less than the range of the set. If δ is less than the range of the set or greater than the spacing between the points, $N(\delta) = \delta^{-0}$ (single point), δ^{-1} (points on a line), and δ^{-2} (an even distribution on the plane). The box dimension is conveniently estimated by dividing the E -dimensional Euclidean space containing the set into a grid of boxes of size δ_{\max}^E , and counting the number of boxes $N_{\text{box}}(\delta)$ that are not empty.

The length spanned by a line is also measured and since the line is straight, the dimension thus measured is termed the ruler dimension D_R (to be discussed later). If the measured line is fractal, $L(\delta)$ depends on the length characteristic δ , such that $L(\delta) \approx \delta^{(1-D_R)}$. From the previous definition, $D_R = 1$ if the line is smooth, suggesting $L(\delta)$ is constant for all values of δ , and the length of the line is simply $L(\delta) = \delta N(\delta)$. Based on the discussion on dimension, the fractal dimension is unique, and is simply the dimension explained by the most advantageous way to measure it. This, of course, depends on whether or not we measure a one- or two-dimensional plane. It should be noted here that the box dimension is widely accepted, easily used, and is close to the Hausdorff–Besicovitch fractal dimension, mathematically expressed as

$$M_d = \sum \gamma(d)\delta^d = \gamma(d)N(\delta)\delta^d \xrightarrow{\delta \rightarrow 0} \begin{cases} 0 & \text{for } d > D \\ \text{or} \\ \infty & \text{for } d < D \end{cases} \tag{16.58}$$

where M_d is the d -measure of the set and its value for $d = D$ is often finite, but can also be zero or infinite; $\gamma(d)\delta^d$ is a test function $h(\delta)$, and represents a disk, line, cube, square, or ball. If we have a line, cube, or square, the geometrical factor $\gamma(d) = 1$; $\gamma = \pi/6$ for spheres; and $\pi/4$ for disks. The position of the jump in M_d as a function of d is most important. Also, D in the Hausdorff–Besicovitch dimension is a local property because it measures properties of sets of points, in the limit of a decreasing size δ of the test function used to cover the set; D can therefore, depend on position. However, the Hausdorff–Besicovitch dimension is very difficult to compute for a real-life data set and therefore is of little functional use.

16.4 FRACTAL CONSTRUCTION

Fractal Dimension $0 < D < 1$

Both triadic and quadric cantor dust are examples of fractals of dimension $0 < D < 1$, and are also classified as “exactly self-similar” fractals. Exactly self-similar fractals are sets of points that cluster on a line segment, but do not fill that line segment. To construct the triadic cantor dust, we begin with a line segment of unit length ($n = 0$) and divide it into three equal parts. Now, eliminate the middle (central) third ($n = 1$) and repeat this procedure on the two remaining line segments ($n = 2$); repeat the procedure as many times as desired. It is readily apparent that as $n \rightarrow \infty$, the points cluster on the line segment of unit length in which their dimension is greater than a single point, but less than the dimension of the line segment of unit length. The dimension of the triadic cantor dust is

$$D_s = \frac{\log [N(\delta)]}{\log (1/\delta)} = \frac{\log (2)}{\log (3)} = 0.631 \quad (16.59)$$

Here (as discussed earlier), the initiator is a line segment of unit length, the generator consists of two segments ($\delta = 1/3$) that contain $N(\delta) = 2$ copies of the original. The construction of the quadric cantor dust is similar except that the two segments of the line are $\delta = 1/4$, which yields a $D_s = \log (2)/\log (4) = 0.500$.

In the above example, the central portion of the line segment was removed to create exactly self-similar fractals. By removing one-third of the line segment at random, the recursive construction is also randomized, and scale invariant—or statistically self-similar—fractals are constructed. By use of power-law distribution, a sequence of power-law-distributed gaps is generated. If a speck of dust is placed between each of these gaps, a cantor dust of any fractal dimension can be constructed. For cantor dusts, it is important to remember that the fractal dimension is really a measure of the degree of clustering on the line segment. Thus, if the dust is highly spread out, the fractal dimension is near zero, and as the dust tends to cluster along the line segment, the fractal dimension is near 1.

Fractal Dimension $1 < D < 2$

Earlier, we discussed the von Koch curves and how finer and finer detail is recursively added, an example of fractal construction beginning with an object of a lower dimension than the fractal dimension sought. Another way to achieve a desired fractal dimension is to begin with an object (volume, surface, etc.) that has a larger dimension than the fractal dimension desired, and cut holes in it until the desired dimension is obtained. Examples of this type of fractal dimension are the “Sierpinski gasket” and “Sierpinski carpet.” The Sierpinski gasket is recursively generated using an equilateral triangle as a generator, and an initiator that has the central portion of this equilateral triangle removed. Sierpinski curves arrive from an

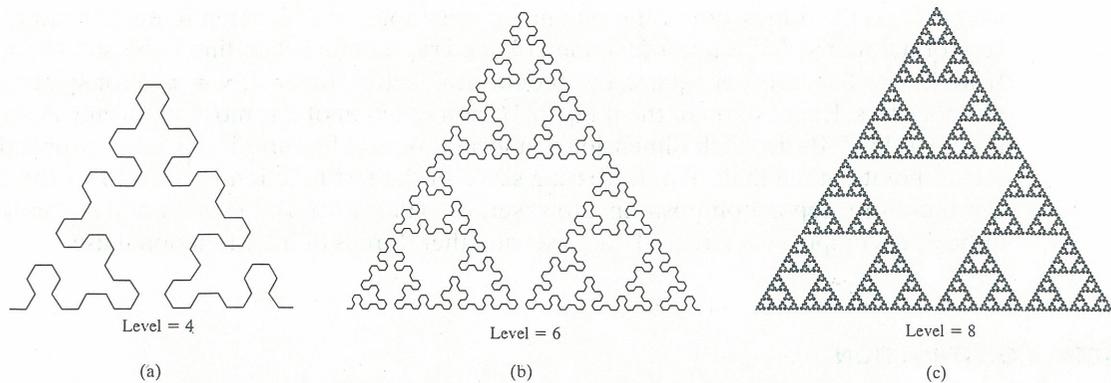


Figure 16.11 Various levels of a Sierpinski gasket (data from Barnsley et al. 1988)

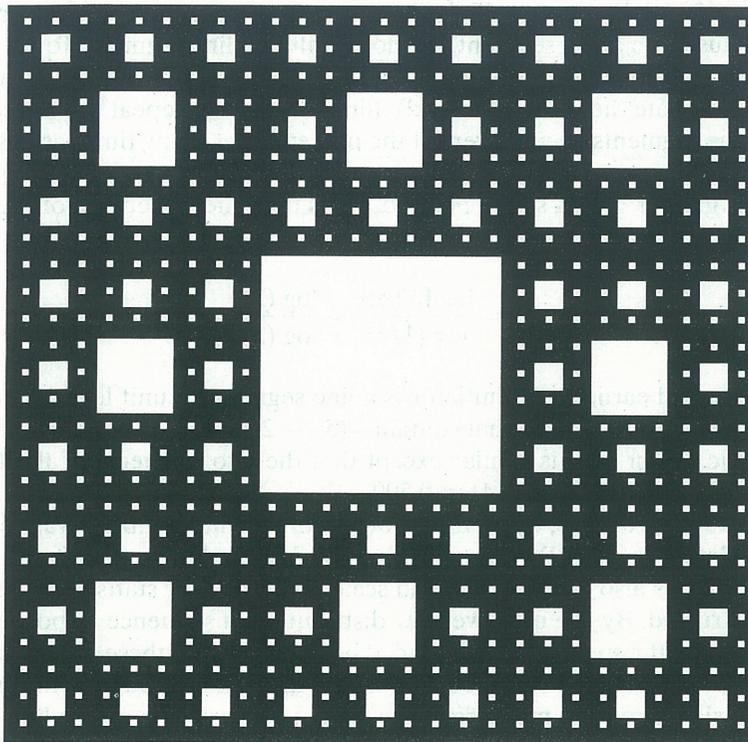


Figure 16.12 A Sierpinski carpet (data from Barnsley et al. 1988)

infinite number of generations of prefractals to leave a fractal curve; this fractal curve was obtained from the Sierpinski gasket (figure 16.11).

Sierpinski carpets have been used as models for the study of soil-water retention, Stokes' flow, pore geometry, pore-volume distribution, and other physical phenomena in soil. The relations to flow in soil is discussed more fully later. A simple way to obtain a Sierpinski carpet is to take a whole surface and randomly cut holes in this surface, where the holes cut have a power-law size distribution. For this case, the initiator contains $N(\delta) = 3$ duplicates of the original line segment of size $\delta = 1/2$ thus, the similarity dimension is $D_s = \log(3)/\log(2) = 1.59$. Unlike the Sierpinski gasket, the Sierpinski carpet (see figure 16.12) is recursively constructed beginning with a unit square, in which the generator is a square with the central square removed. The initiator contains $N(\delta) = 8$ duplicates of the original line

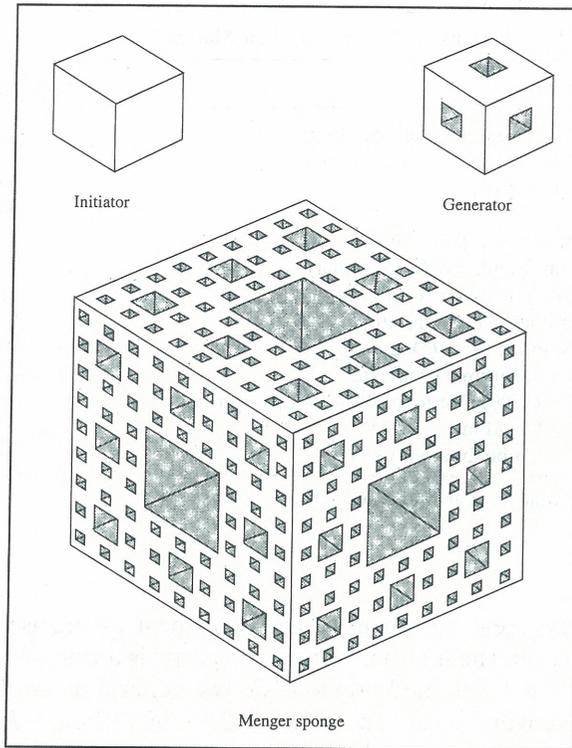


Figure 16.13 A Menger sponge shown with its initiator and generator (data from Barnsley et al. 1988)

segment of size $\delta = 1/3$. As a result, the similarity dimension $D_s = \log(8)/\log(3) = 1.893$. Both of these are examples of exactly self-similar fractals, and the pattern of distribution for holes within each type follows a power-law distribution.

Fractal Dimension $2 < D < 3$

An example of this fractal type is the Menger sponge (see figure 16.13), which is a volumetric analog of the Sierpinski carpet. As we might suspect, the initiator is a unit cube divided into 27 smaller cubes to form the generator, and then 6 of the small cubes are removed from the center of each face, as well as the small cube at the center. Each of these smaller cubes is of size $\delta = 1/3$. The generator possesses $N(\delta) = 20$ duplicates of the original segment. The similarity dimension is $D_s = \log(20)/\log(3) = 2.727$; fractal dimensions can—and will—vary considerably for different materials.

16.5 SELF-SIMILAR FRACTALS: ESTIMATING THE FRACTAL DIMENSION

Now we discuss measuring the fractal dimension of self-similar fractals. As with self-affine fractals, we measure a characteristic of a data set that is related through some sort of power law to a length scale and as before, the results are plotted on a log-log scale. It is important to remember that more than 200–300 discrete measurements are needed at a variety of length scales, in order to make certain the data set has scale-invariant characteristics. From a physical viewpoint such as water flow or contaminant transport, it is important to obtain the fractal dimension because it can correspond to the effects of various physical processes that occur in soil. A partial list of fractal dimensions for various materials is given in table 16.1.

TABLE 16.1 Fractal Dimensions for Various Geologic Materials

Description of material	Fractal dimension
Upper Columbus dolomitic rock (Bellevue, Ohio)	2.91
Granitic rock (SHOAL nuclear test site, Nevada)	2.88
Soil (kaolinite, trace halloysite)	2.92
Porous silicic acid	2.94
Coal-mine dust from western Pennsylvania	2.52
Mosheim high calcium (Stephens City, Virginia)	2.63
Niagara (Guelph) dolomite (Woodville, Ohio)	2.58
Soil (mainly feldspar quartz and limonite)	2.29
Aerosil—nonporous fumed silica (Degussa)	2.02
Madagascar quartz from thermal syndicate	2.14
Graphite—Vulcan 3G (2700) (National Physical Laboratory, Teddington, United Kingdom)	2.07
Iceland spar, massive (Chihuahua, Mexico)	2.16

Source: Data from Avnir, Farin, and Pfeifer (1984).

Grid Dimension

This measuring technique is so termed because boxes of linear size δ are used to measure a set; these boxes usually form a grid. The number of boxes required is termed $N(\delta)$ and the dimension has the relation $N(\delta) = \delta^{-D}$ (defined previously). The general technique is to count the boxes required of size δ , to cover the set of a range of values of δ . Then, $-D$ (the slope) is obtained by plotting $\log_{10} N(\delta)$ versus $\log_{10} \delta^{-1}$. As with the previous measurement techniques, if the set is fractal, the plot is a straight line with a negative slope equal to $-D$. For evenly spaced points on a log-log plot, the size of the box chosen follows a geometric progression.

Point Dimension

The point dimension—also referred to as the pointwise, cluster, or mass dimension—allows us to count the number of points within a set that has been encompassed by a circle of radius r . In this case, the circle does not include the entire set, but only that part of the set that fits within the defined radius. To imagine this, we draw a set of points in a straight or curving line, and select a circle of radius r to encompass a portion of the points, as in figure 16.14. This can

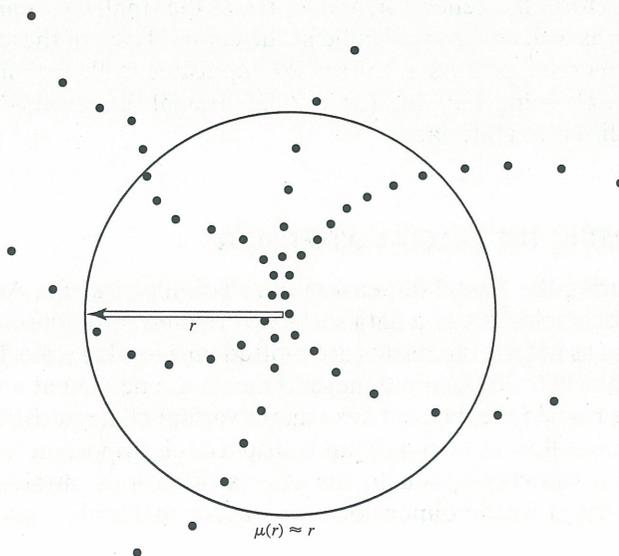


Figure 16.14 Exploded view of a point dimension

also be done on a plane. The mathematical relation is

$$m(r) = \frac{P(r)}{P} \quad (16.60)$$

Since P is the number of points in the entire set, the mass is $m(r)$ within the circle of defined radius r , and the number of points within the circle is $P(r)$. The mass within the circle is proportional to r in one dimension, and r^2 in two dimensions. In real-life soil applications, a beginning measurement point (preferably at the center) is selected and circles of increasing radius are used to measure the mass $m(r)$. The next step is to plot $\log_{10} m(r)$ versus $\log_{10} r$ that follows a straight line if the set is fractal, as described in previous measurement methods. The relation for the mass in this case is $m(r) \approx r^D$, where D is usually written D_p to denote point dimension. This method is particularly applicable to sets that have a radial symmetry—that is, diffusion-limited aggregation (DLA).

Ruler (Divider) Dimension

An excellent example of the “ruler” dimension is the measurement of a coastline, given by Feder (1989) and other authors that describe fractal measurement. If we observe a world map and investigate the coastline of any country, at first glance it appears relatively smooth. However, upon closer inspection the mouths of rivers, bays, and various tributaries and features become more prominent. By the same example, if we use a ruler of length δ , the length of the coastline can be conveniently measured. If the ruler length were large—except for the basic outline of the coast itself—the detail of the various topological features (discussed above) is lost. By decreasing the ruler length δ , more detail of the coastline is apparent. The total length of the coastline is also increased because of this detail. The ruler dimension is normally expressed as D_R and is expressed mathematically as

$$L(\delta) \approx \delta^{1-D_R} \quad (16.61)$$

where $L(\delta)$ is the length of the coastline (or measured) object by the ruler (or divider) of length δ .

If the line is Euclidean, $D_R = 1$, the overall length is independent of δ if δ is sufficiently small compared to the measured object. However, if the ruler fills the space completely, the length of the line is linearly related to the ruler length. Once $L(\delta)$ has been determined by a ruler of length δ , we can plot the $\log_{10} L(\delta)$ versus the $\log_{10} \delta$ if the line is fractal, the plot will follow a straight line as before. The plot has a negative slope equal to $1 - D_R$. We note that the log-log plot of the measured length shows no sign of reaching a fixed value as δ is reduced.

Perimeter–Area Dimension

If we imagine a set of natural, yet geometrical, objects within a two-dimensional plane (e.g., tree leaves floating on the surface of a pond), we can parameterize the leaves as a group of Euclidean objects with area A and perimeter P . We can prove that A and P of each leaf within the group is related by

$$A = \pi r^2 = \frac{P^2}{4\pi} \approx P^2 \quad (16.62)$$

$$P = 2\pi r = r\sqrt{\pi A} \approx \sqrt{A}$$

where r is the radius of each leaf. Also, because r is independent, the proportionality between the area and perimeter is independent of r . If the perimeter of the leaves is fractal, then the relation of A to P is

$$A \approx P^{2/D_{P-A}} \approx P^2 \quad (16.63)$$

$$P \approx [\sqrt{A}]^{D_{P-A}} = A^{D_{P-A}/2}$$

where D_{P-A} is the perimeter–area dimension. Plot $\log_{10} A$ versus $\log_{10} P$ on the vertical versus horizontal axis; if the set is fractal, the slope is equal to $2/D_{P-A}$ in the positive direction. This is accomplished by using δ at a fixed, small finite scale or resolution.

16.6 SELF-AFFINE FRACTALS

The coastline example discussed earlier is statistically self-similar for any given value of the scaling ratio r , as well as for all scaling ratios between some minimum and maximum cutoff values. For these types of fractals, the box-counting method is used to estimate the fractal dimension. In nature, however, various cases of interest are not self-similar. For example, the motion of a Brownian particle has different physical quantities for both position X and time t . As a result, these two quantities do not scale to the same ratio, and the related fractal is termed “self-affine.” The parts of these fractals need to be rescaled, by different ratios in different coordinates, to resemble the original. Mathematically, a bounded set S is self-affine with respect to a ratio vector such that $r = (r_1, \dots, r_E)$ if S is the union of N nonoverlapping subsets S_1, \dots, S_N . Each set has to be congruent (i.e., the set of points S_i is identical to the set of points $r(S)$ after possible translations of the set) to the set $r(S)$, obtained from S by the affine transform defined by r . The affine transform converts a point $x = (x_1, \dots, x_E)$ into new points where the scaling ratios r_1, \dots, r_E are not all equal such that $x' = (r_1 x_1, \dots, r_E x_E)$. In these equations, the subscript E refers to the Euclidean dimension.

Hurst’s Empirical Law (Exponent H)

From a lifetime study of the Nile River, Hurst (1951) invented a new statistical method for analysis termed the “rescaled range analysis” (R/S analysis). After investigating river discharges, mud sediments, and other natural phenomena surrounding the Nile, Hurst used the dimensionless ratio R/S for data analysis, where R is the range and S is the standard deviation. During years of data collection and analysis, Hurst developed the following mathematical relation:

$$\frac{R}{S} = \left(\frac{\tau}{2}\right)^H \tag{16.64}$$

where H is the Hurst exponent and τ is the number of years (number of tree rings, number of times a coin is tossed, and so on). Actually, τ can be thought of as the number of observations. If we consider a self-affine fractal such as the Devil’s staircase (Feder 1989; see figure 16.15), we obtain an exact copy of the original only by scaling portions of the curve by different factors, such as $r_x = 3$ in the x direction and $r_y = 2$ in the y direction. These are the

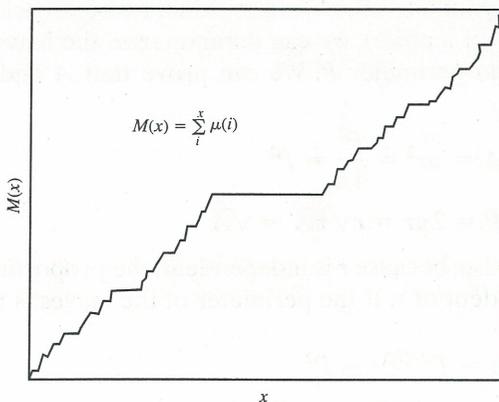


Figure 16.15 Illustration of the “Devil’s Staircase” (data from Feder 1989)

rescaling factors that can be rewritten to obtain the Hurst exponent H , such that $r_x = r$ and $r_y = r^H$. Solving for H yields

$$H = \frac{\log(r_y)}{\log(r_x)} \quad (16.65)$$

which is the general form of the equation and for which H in the Devil's staircase example is 0.631—that is, $\log(2)/\log(3)$. Also, because r_y is written as a power of r_x , the relation is scale invariant (i.e., independent of length). The standard deviation S is given by

$$S = \left(\frac{1}{\tau} \sum_{t=1}^{\tau} [\xi(t) - (\xi)_{\tau}]^2 \right)^{1/2} \quad (16.66)$$

where τ is the window (period) of time over which the observations are taken, $\xi(t)$ is the specific time and value of an observation, and ξ_{τ} is the average (or mean) value of observations for the time over which all observations are taken. From data gathered by Hurst, S is about 0.09 and is (as a rule) symmetrically distributed about a mean of 0.73. Hurst also shows that for numerous natural phenomena, $H > 0.5$. For statistically independent processes with finite variances in the absence of long-term statistical dependence, $R/S = (\pi\tau)^{1/2}$, it is also asymptotically proportional to $\tau^{1/2}$.

Brownian Motion and Random Walks in One Dimension

Consider the tossing of a coin. Each time the coin is tossed, there is a probability $p = 0.5$ of that coin being a head or a tail. Thus, if we assign a value of $+1$ for heads and -1 for tails, we can obtain a record of the trace of the compilation of tosses made (see figure 16.16). This is analogous to a particle moving on a line (assume the line is the x -axis), on which the particle jumps a step-length of $+\xi$ or $-\xi$ for a given time interval t (time interval in second, minute, hour; sometimes referred to as “collision time” in random walks). Such a stepping motion is referred to as a “random walk” (or Brownian motion) just as can be seen observing floating dust particles against a dark background. In one dimension, the random walk takes place only in the vertical coordinate. Also, the displacement of the individual particle in the given time interval is independent of the displacement of the same particle during another time interval. By consideration of itself, Brownian motion is self-similar; however, when considering particle position as a function of t , Brownian motion is self-affine. In the latter case, the Hurst exponent $H = 0.50$.

Considering a Gaussian distribution of zero mean, Brownian motion is a random walk in which the length of each step has nothing to do with the length of a neighboring step.

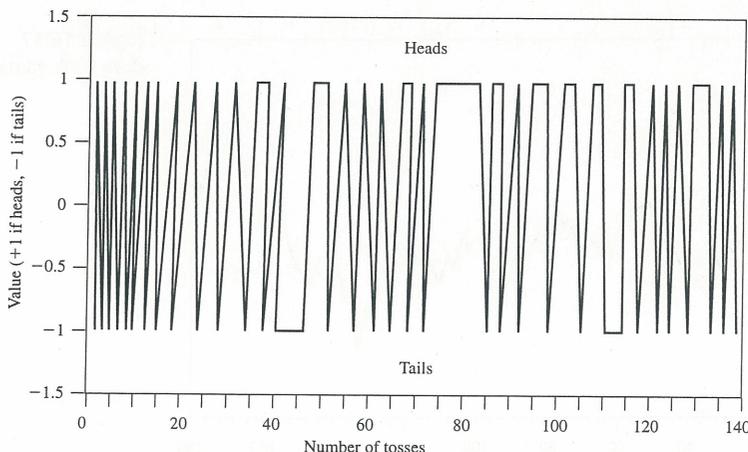


Figure 16.16 A compile trace of coins tosses (data from Feder 1989)

Consequently, the normal probability distribution (npd) is given by

$$p(\ell, t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{\ell^2}{4Dt}\right) \quad (16.67)$$

where D is the diffusion coefficient given by the Einstein relation

$$D = \frac{1}{2t} \langle \ell^2 \rangle \quad (16.68)$$

The parameter $\langle \ell^2 \rangle$ is the mean-square jump distance. The Einstein relation is only valid under general conditions—that is, jumps do not occur at regular intervals, and when the npd for ξ is continuous, discrete, or has some arbitrary shape. The step-length ξ has to be chosen at random for any time interval t , such that the probability of finding ξ from ξ to $\xi + d\xi$ is $p(\xi, t) d\xi$. This yields a set of independent, random variables (Gaussian), for which the variance is

$$\langle \ell^2 \rangle = \int_{-\infty}^{\infty} \ell^2 p(\ell, t) d\ell = 2Dt \quad (16.69)$$

By letting ξ become $\xi/(2Dt)^{1/2}$, a normalized Gaussian random process is obtained. In the normalized process, ξ has a zero average with a variance $\langle \xi^2 \rangle = 1$. An example of a Gaussian random walk is shown in figure 16.17.

Scaling Properties of Random Walks in One Dimension

Consider sitting on a chair in a sunny room, observing floating dust particles. As sunlight reflects off each particle, the particle is clearly seen against shadows within the room or against a dark background. By focusing on only one of the dust particles, it appears to float effortlessly in a constant motion. As a result, we cannot clearly and finitely resolve the position of the particle; we can see it only at intervals bt , where b represents the first, third, or some other time step. Suppose we see the second time step of a particle's position; any increment ξ of the particle position is now a sum of the two steps (i.e., ξ_1 and ξ_2). Regardless of the number of b collision (or microscopic) time steps, the change in particle position is an independent Gaussian process with $\langle \xi \rangle = 0$ and a variance of $\langle \xi^2 \rangle = 2Dt$, where $t = b\tau$.

Because Brownian motion is self-affine, transformation of the normal probability distribution is accomplished by replacing $\tau = b\tau$ and $\xi = b^{1/2}\xi$; thus, the length scale is changed by a factor of $b^{1/2}$ and the time scale by a factor of b , which yields a scaling relation of

$$p\{\xi = \sqrt{b}\xi, \tau = b\tau\} = -\sqrt{b} p(\xi, \tau) \quad (16.70)$$

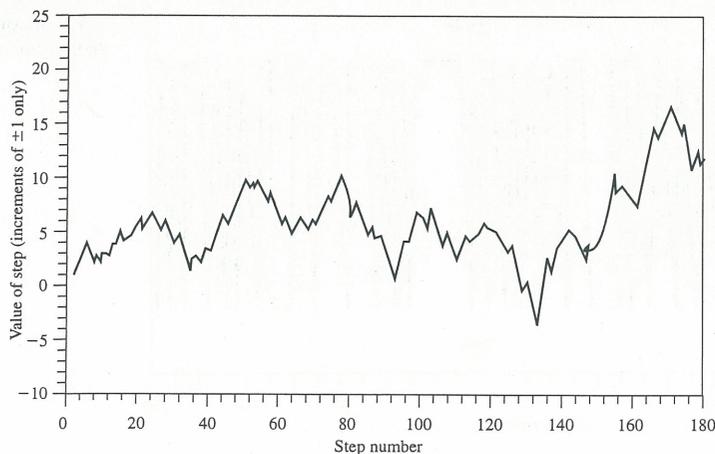


Figure 16.17 A Gaussian random walk (data from Feder 1989)

Hence, the scaled property appears the same, as if not scaled. Consequently, we effectively scaled and changed the resolution, and the Brownian record looks the same (i.e., it is scale invariant). Also, $b^{-(1/2)}$ ensures that the normal probability distribution is appropriately normalized.

Fractional Brownian Motion: Mathematical Models

Fractional Brownian motion, fBm, has become one of the most useful mathematical models for generating random fractals like those found in nature. This includes self-similar fractal landscapes, clouds, and mountainous terrain. Figure 16.18 shows fBm traces, usually denoted $V_H t$, plotted as a single-valued function consisting of only one variable t (time). The traces shown in figure 16.18 represent the point of location or differences of fBm between successive intervals. As we suspect, the scaling behavior of each trace is characterized by the Hurst exponent H , with a range $0 < H < 1$. The closer H approaches to 1, the smoother the trace, and as H gets closer to 0, the rougher the trace. H relates any change in quantity V ($\Delta V = V(t_2) - V(t_1)$), to a difference in time ($\Delta t = t_2 - t_1$) by the scaling law $\Delta V \propto \Delta t^H$. The sum of independent steps for random walks leads to a variation that scales as the square root of the sum of steps; in this instance $H = 1/2$ corresponds to a trace of fBm. The Gaussian distribution of fBm has a variance

$$\langle |V_H(t_2) - V_H(t_1)|^2 \rangle \propto |t_2 - t_1|^{2H} \quad (16.71)$$

where $\langle \rangle$ and $\langle \rangle$ are grouping averages consisting of many samples of $V_H(t)$ and H has the value $0 < H < 1$. The mean-square steps depend only on Δt ; each t is statistically equivalent. It is important to note that fBm is not differentiable, even though $V_H(t)$ is continuous. Constructs (based on averages $V_H(t)$) are developed to give meaning to a derivative of fBm. The derivative of normal Brownian motion is $H = 1/2$ and for $H > 1/2$ there is a positive correlation for increments of $V_H(t)$ and its derivative fractional-Gaussian noise, while for $H < 1/2$, there is a negative correlation.

A statistical scaling behavior for $V_H(t)$ is shown when the time scale t is altered by a factor τ , such that the steps ΔV_H are altered by a factor τ^H . This scaling behavior is given by

$$\langle \Delta V_H(\tau t)^2 \rangle \propto \tau^{2H} \langle \Delta V_H(t)^2 \rangle \quad (16.72)$$

Here, the t coordinate has a special status. While each t corresponds to only one value of V_H , any specific value for V_H can occur at multiple t 's. This nonuniform scaling is termed self-affine.

Global Dimension of Self-Affine Fractals

The global dimension of self-affine fractals is $D = 1$, and is equal to the fractals' topological dimension; the local dimension is fractal and $D = 1/H$. The local dimension is often termed

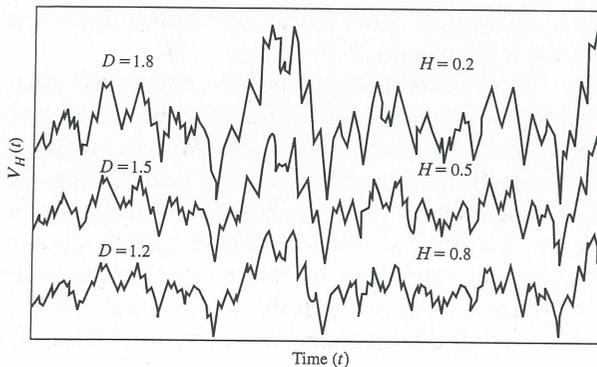


Figure 16.18 A trace of fractional Brownian motion (fBm); the point of location or differences of fBm between successive intervals (data from Peitgen and Saupe 1988)

the “latent fractal” dimension because it is related to the fractal dimension associated with the trail of a specific Brownian particle. When $D = 1$, the fractal trace resembles a straight line; this infers that there is no vertical amplitude. The amplitude of the trace is also known as the “cross-over” scale, and it characterizes the overall roughness of the fractal surface. Thus, we can obtain two profiles with a similar fractal dimension, but totally different surface roughness. An example of this is the similar fractal dimensions of a concrete runway and mountainous terrain (Sayles and Thomas 1978), however, the cross-over scales for each of these is completely different. As a result of these parameters, two observations are made. First, there is a ratio of the vertical range (R) over the horizontal range L . The ratio can be written as $R(L)/L = w^H - 1$, where w is a window length (size) in which measurements are taken and H is the Hurst exponent. As the length L of the trace becomes much larger than the cross-over scale, any irregularity within the trace is minimal or if $H < 1$, the ratio approaches zero as L approaches infinity. Second, the self-affine record has a global value of the fractal dimension $D = 1$ —that is, globally, a self-affine record is not fractal. As a result, it is important to measure the local dimension.

Measurement of Self-Affine Fractals

In discussing the global dimension of self-affine fractals, we have indicated that the ratio between the vertical and horizontal ranges is scale dependent (i.e., each varies with changes in scale). Also of note, the techniques used to measure self-similar fractals are inappropriate for self-affine fractals. For example, if the length of the ruler used to measure a self-affine fractal exceeds the cross-over scale discussed earlier, that measured length normally remains constant and the resulting dimension is always $D = 1$. Thus, since we want to measure the local dimension and not the global dimension, this is inadequate. Additionally, if we have a large scale in comparison to the cross-over scale, D is also one. In contrast to using boxes, grids, or rulers, as with self-similar fractals, we have to measure the change in vertical range (roughness) over different horizontal scales that yield an estimate of H (Hurst exponent). If the Hurst exponent is obtained, it can be converted to a fractal dimension, as proven earlier. To obtain this measurement, we need a variety of length scales that vary on an order of magnitude and a few hundred samples.

Roughness/length This method of measurement is used when there is no constant sampling interval. However, we assume that each sampling interval (window) has about 5–10 samples. Instead of using the vertical range, we calculate the standard deviation of roughness data for windows of size w ; this is related to the Hurst exponent by $\sigma(w) \approx w^H$. The entire length is measured by a series of windows and for each window, the roughness is determined after subtracting any local linear trends. As a final estimate, the average roughness for each window $\langle \sigma(w) \rangle$ is determined. For a number (series) of window lengths, plot $\log_{10} \langle \sigma(w) \rangle$ versus $\log_{10} w$. For a self-affine trace this plot is a straight line. The slope of this line is the Hurst exponent and is determined by dividing the horizontal range by the vertical range—that is, $H = \log(x)/\log(y)$. Once H is calculated, the fractal dimension D for a self-affine trace is given by $D = 2 - H$, and for a surface is given by $D = 3 - H$.

Rescaled range (R/S) To use this method, the data measured must have a constant sampling interval. This is because the expected difference between successive values of $y_1 \dots y_n$ is a function of the distance between each y value. We already have the equation for R/S from the beginning of this section, so it is not repeated here. Suffice it to say that R represents the range of values extracted over the y interval that is measured as a line segment, connecting the first point to the last point within a window (i.e., with respect to a trend within the window); S is simply the standard deviation. In this method, R is the average of a number of values. As with the roughness/length method, if the trace is self-affine, the plot follows a straight line. The slope of this line is the Hurst exponent H . Also, the fractal dimension D for

both a trace and a surface are the same as for the roughness/length method, except that the R/S method has to be employed to obtain D . Thus, we write $D_{R/S}$.

Variograms This measuring method can also be applied to a series that does not have a constant sampling interval, provided that points are selected such that $\Delta\tau$ is a small percentage of τ . By squaring the difference between two y values in a trace, separated by the interval τ , we have a variogram, which is related to the autocovariance function (Journal and Huijbregts 1978). This is written as

$$\gamma(h) = \frac{1}{N} \sum_{\alpha=1}^N (z(x_{\alpha} + h_{\alpha\beta}) - z(x_{\alpha}))^2 \quad (16.73)$$

where $\gamma(h)$ is the dissimilarity in the two values (squared difference), $z(x)$ is the regionalized variable at points γ and β , and N is number of points sampled. Dividing equation 16.73 by 2 (commonly seen in the literature) yields a semi-variogram. Previously, we listed the standard deviation in relation to fractional Brownian motion as $\sigma(\tau) = \tau^H$; if we now consider a normal distribution with a squared difference, we can write $V(\tau) \approx \tau^{2H}$. To obtain H , the $\log_{10} V(\tau)$ is plotted versus the $\log_{10} \tau$. Again, the slope has to be a straight line if the fractal is self-affine and represents $2H$ because of squaring. Thus, for true H , divide the slope by two. The fractal dimension D_v for this method for a trace and a surface is the same as the previous. Through measurement, an average trend appears that has to be subtracted before estimating range or roughness. The variogram accurately measures the average squared value of the trend. It should be noted that the average trend needs to be subtracted before range or roughness is estimated, to prevent error from creeping into the calculations.

Power-spectral density As a function decomposes into a sine or cosine (harmonic) function, it is termed the “power-spectral density,” PSD: the PSD is the squared amplitude of a harmonic function. To obtain the fractal dimension, the power spectrum $P(k)$ must be determined; $\log_{10} P(k)$ is then plotted versus $\log_{10} k$. The harmonic function is expressed as wavelength λ , frequency f , or wavenumber k . Both the frequency and wavenumber is $1/x$, where x is either selected as time or length. The relation between all three is $k = 2\pi f = 1/2\pi\lambda$. As a consequence, the power spectrum for self-affine fractals follows a power law (scale-invariant); the exponent is equal to $-\beta$. As with the other methods already discussed, if the trace is self-affine, the plot follows a straight line with a negative $-\beta$. The reader should recall that for a self-affine trace, $D = 2 - H$. Also, the method can be extended to two-dimensional surfaces, by a measure of $V(w)$ as the average squared distance between points. The points are separated by some distance w in all directions. Thus, the appropriate fractal dimension of the surface could be expressed as $D = 3 - H$.

The scale-invariant form of power spectrum of a self-affine trace and surface can be shown with Parseval’s theorem, relating the spectral density $P(k)$ of a function $y(x)$ to its variance σ^2 :

$$\int_{-\infty}^{\infty} y(x)^2 dx = \sigma^2 = \int_{-\infty}^{\infty} P(k) dk = 2 \int_{-\infty}^{\infty} P(k) dk \quad (16.74)$$

where $y(x)$ has zero mean and $P(k)$ is an always-positive function symmetric about zero. Parseval’s theorem illustrates that the power spectral density $P(k)$ quantifies how the total variance in a function is partitioned in components of varying wave number, k . Because the variance of a self-affine function depends on the length of the window w , and the length scale is proportional to wavelength λ , the inverse of equation 16.74 is expressed as

$$(k) = \frac{d}{dk} \sigma^2(w) \approx \frac{d}{dk} w^{2H} \approx \frac{d}{dk} \lambda^{2H} \approx \frac{d}{dk} k^{-2H} \approx k^{-2H} \approx k^{-(2H+1)} = k \quad (16.75)$$

The power-spectral density for a self-affine trace, as seen in equation 16.75, takes the form of

a power law with an exponent equal to $-\beta$. Using equation 16.75, β is related to the fractal dimension D to obtain

$$D = \frac{5 - \beta}{2} \quad (16.76)$$

Since we are discussing a self-affine surface, $P(k)$ is a function of wave numbers in k_x and k_y , respectively. For two-dimensional surfaces, $P(k)$ is a function of wave numbers in both the x and y direction; however, by defining k as independent of direction, $P(k)$ is a function of only one variable. This is defined by

$$k = \sqrt{k_x^2 + k_y^2} \quad (16.77)$$

In this instance, the fractal dimension $D = (8 - \beta)/2$. This method is the most difficult of all that we have discussed for obtaining the fractal dimension, although there are some who believe it the most cited in the literature. One reason for this: using the expression for k —taken as a squared-Fourier transform of a finite series—is a poor estimate of $P(k)$. This squared transform is termed a “periodogram,” whose estimate of power at various frequencies is very noisy, the amplitude of the noise proportional to the function’s spectral power. If the amplitude is filtered to make it more like the amplitude of “white noise” (for typical Brownian increments $H = 1/2$; the noise in this case is an independent Gaussian process normally called white noise), a more reliable estimate of the fractal dimension is obtained. In summary, we have presented four methods for measuring self-affine fractals: **(1)** roughness/length, for use when there is no constant sampling interval; **(2)** rescaled range, to be used when data measured have a constant sampling interval; **(3)** variograms, a method to be applied to a series that does not have a constant sampling interval; and **(4)** power-spectral density, a statistical tool commonly used to study the frequency content of signals. The first two methods are the easiest to use; however, before attempting the latter two, the reader should consult current literature on the subject. Because there is so much literature, we have conducted only a brief discussion here, and the reader is referred to the References section at the back of the book for a more complete list of articles on the subject.

Hyperbolic distribution The pdf for a random variable written in terms of a hyperbolic distribution (also called a power law) is expressed as $p(x) = bCx^{-(b+1)}$, where b is some power and C is a constant. This expression is only true for $x > 0$ and $b > 0$. The modeling of random variables using a hyperbolic distribution is only useful for positive random variables, and is most effective where the number of smaller members (within the group being sampled) is greater than the number of larger members. Here, b characterizes the hyperbolic distribution since C is only a constant and not very important, except that it depends on the minimum value of x chosen. Essentially, the hyperbolic distribution has all the characteristics of scale invariance. In such cases, the exponent b is invariant under multiplication. Unlike the previously discussed distributions, there is no scale length associated with hyperbolic distributions. Thus, hyperbolic distributions are scale invariant and any process that creates a variable in this manner must also be scale invariant.

16.7 VISCOUS FINGERING AND DIFFUSION-LIMITED AGGREGATION

Multifractals (Feder 1988) are shapes and measures that require more than one dimension, and have received broad attention in the study of such dynamical systems as Poincare maps and electric-field strength in aggregation problems. Consequently, the measurement of multifractals is related to the study and distribution of physical and other quantities on the surface of a sphere or an ordinary plane. Additionally, the support itself can be a fractal. Fractals provide a general language for the classification of various pathological or indescribable shapes (“animals”) encountered in the natural sciences. The various ways in which matter

condenses on the microscopic scale seem to generate fractals; this is an example of percolation, often confused with diffusion (Feder 1988; Peitgen and Saupe 1988). An example of diffusion is the random motion of dye particles mixed in a solution. In contrast, percolation is displayed when the randomness is attached to the medium itself. For example, consider water beading on the hood of an automobile. Initially it is several atomic layers thick but rather than a uniform coverage, the water beads, due to surface tension. Analysis of the connected clusters (droplets) shows irregular branching structures of finite size. As the amount of water increases, the clusters increase in size, and eventually connect across a certain sample length. The unique characteristic associated with percolation is that of a percolation threshold, p_c . If the rate of buildup is below p_c , then the spreading of fluid is confined to a usually small, finite area. Assuming that a point source is constant and that water is then applied, is the resulting contaminant going to be contained locally within a droplet, or is it going to spread across the hood and connect to other beading droplets, causing a large puddle that can connect to other large puddles? The transport of immiscible fluids (e.g., oil and water) is also a good example that can create fingering, invasion percolation, or preferential flow.

Diffusion-Limited Aggregation (DLA)

DLA is a simple model that reproduces many natural shapes, such as electrochemical deposition, electrostatic discharge and—for purposes considered here—fluid–fluid displacement. It is easily implemented on a computer, with resulting structures resembling those of root distributions within a soil profile. The DLA process is one in which a particle or monomer diffuses in the “random walk” process (see section 16.6). Imagine a solution of randomly mobile particles; DLA begins with a single fixed, sticky particle at an origin point within this solution. Each of the mobile particles moves on its own random path, one random unit at each time step. At some point, a random particle finds itself next to the origin point and “sticks” to it. This point (beginning from one site) slowly grows into a cluster with the addition of other random particles, to complete the random-walk process. A structure much like that of the distribution of roots within a medium forms with many branches (see figure 16.19). The open branches do not fill, because a particle trying to reach the innermost point of the branch invariably contacts one of the sticky sides first. The cluster thus formed is an example of a

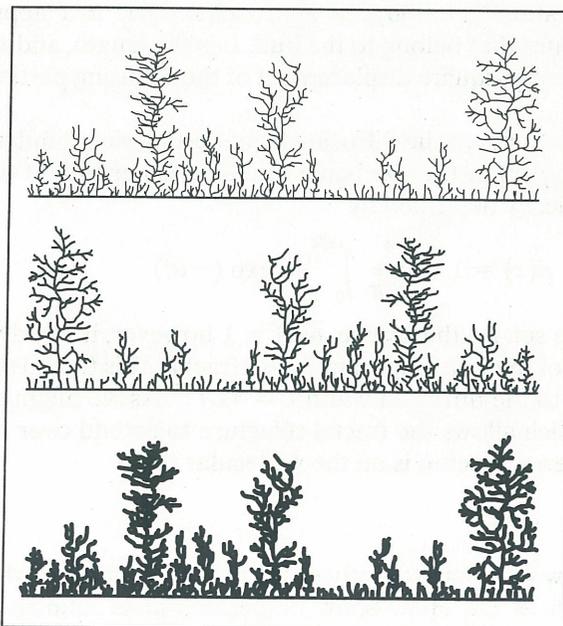


Figure 16.19 A schematic of diffusion-limited aggregation (DLA). The formed clusters are termed “percolation transition,” and are normally fractal in nature (data from Feder 1989)

“percolation transition,” and is usually fractal in nature. The cluster is both porous and random but also, the cluster has to exhibit decreasing density upon increase in size described by

$$\rho(r) \approx R_o^{-D} r^{D-E} \quad (16.78)$$

where ρ is the cluster density, r is cluster radius, R_o is particle radius (particles forming cluster and thus, the characteristic length scale), D is the fractal dimension, and E is the Euclidean embedding dimension (i.e., $D < E$). For a fractal cluster, the density decreases with distance from the point of origin.

Fractal Diffusion Fronts

If we apply methylene blue dye to the top of a sandy soil and add water very slowly in small amounts, a diffusion front begins to appear. Upon close examination, the front appears like the edge of a coastline due to adsorption of the dye to the sand particles. This is another example of percolation. The structure of the diffusion front is fractal and is related to the hull of percolation clusters (Sapoval, Rosso, and Gouyet 1985), but the diffusion process is not fractal. The surface of the diffusion front is called the hull and is a finite fraction of the adsorption sites involved with a fractal dimension. As the particles of dye move from the source, diffusion D is described by the Einstein relation given in equation 16.68 (the particle moves to a neighboring site a distance a , and time step τ). This means that the displacement of a diffusing particle in the x (perpendicular) and y (parallel) directions away from the source are independent of each other.

The position of the hull is given by $p(x_h) = p_c$. Also, if the diffusion has a width of $2L$ instead of L (L is a length over which data is measured), the number of occupied sites belonging to the hull is also doubled, which makes a view of the hull one-dimensional. The hull is a self-similar fractal up to length scales that equal the hull's width. The fractal dimension of the hull can be written as

$$D_h = \frac{1 + v}{v} = 1.75 \quad (16.79)$$

where $v = 4/3$ (controls the divergence of the correlation length ξ at p_c). This was proposed by Sapoval, Rosso, and Gouyet (1985) and proved to be correct by Saleur and Duplantier (1987). This is because the ratio $M_h(L/\ell)\sigma_h/L\ell$ approaches 0.441 as ℓ approaches infinity, where M_h is the number of sites that belong to the hull, L is the length, and ℓ is the diffusion distance defined as the root mean-square displacement of the diffusing particle from its starting point.

At any time step, the particles at the diffusing front occupy sites similar to the percolation process. The probability $p(x)$ of the site being occupied is dependent on distance from the source, and is mathematically described by

$$p(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^{x/\ell} du \exp(-u^2) \quad (16.80)$$

where u is an element of the set. At the source, $p(x) = 1$ however, it rapidly diminishes for $x > 1$. The internal structure of the diffusing front is also fractal. This fractal structure extends over distances proportional to the diffusion width $\ell = 4Dt^{0.5}$. As we might suspect, this distance diverges with time, which allows the fractal structure to extend over macroscopic distances, even in instances where diffusion is on the molecular scale.

Viscous Fingering in Soil

The principles of viscous flow in soil are exactly similar to the principles which control flow in viscous-flow analogs such as the Hele-Shaw or parallel-plate analogs. Both are well-

known devices used for two-dimensional ground-water investigations. The viscous-flow analog is based on the similarity of differential equations that govern flow of a viscous fluid in the narrow space between two parallel plates and those that govern saturated flow in soil. This methodology has been used in investigations of seepage through earthen dams, artificial recharge, drainage, oil production in reservoirs, and in other investigations. The Navier-Stokes equation for a viscous incompressible fluid is given as

$$\frac{DV_x}{Dt} = f_x - \frac{1}{\rho} \frac{\partial p}{\partial x} - \nu \nabla^2 V_x \quad (16.81)$$

where DV_x/Dt represents the hydrodynamic derivative, V_x is the velocity component in the x direction, ν is the kinematic viscosity, ρ is the fluid density, f_x is the component of external force per unit mass acting on the liquid, and p is the pressure. Because the fluid flows through a narrow vertical space of width b , $V_y = 0$. For fluids that have very high viscosities or very slow (creeping) motions, viscous forces are much greater than inertial forces. Thus, assuming this type of flow takes place between the parallel plates, the left-hand side of equation 16.81 (the inertial term) can be neglected. Consequently, the only active-body force is gravity with potential gz . Thus, $f_x = -\partial(gz)/\partial x = 0$; force in the y direction is the same, we simply substitute the x component with the y component—however, $f_z = -\partial(gz)/\partial z = -g$. Since b is narrow but also because the fluid adheres to the parallel plates, velocity gradients in the y direction are considerably larger than those in the x or z directions. As a result, we can neglect $\partial V_x/\partial x$, $\partial^2 V_x/\partial x^2$, $\partial V_z/\partial x$, $\partial^2 V_z/\partial x^2$, when compared with $\partial V_x/\partial y$, $\partial^2 V_x/\partial y^2$, $\partial V_z/\partial y$, $\partial^2 V_z/\partial y^2$. Thus, equation 16.81 becomes

$$\frac{\partial(p + \rho gz)}{\partial x} = \mu \frac{\partial^2 V_x}{\partial y^2} \quad (16.82)$$

The parameter $(p + \rho gz)$ remains constant in the y direction and thus is equal to zero. Rearranging equation 16.82 for velocity we obtain

$$V_x = -\frac{k_x}{\eta} \nabla(p + \rho gz) \quad (16.83)$$

where k is the permeability of the medium and η is the dynamic viscosity of the fluid. The term k/η is referred to as the mobility—thus, the velocity $V = -M(p + \rho gz)$. For a horizontal position of the parallel plates, $k = b^2/12$ (only for the Hele-Shaw cell; for soil, k is the actual measured permeability). In terms of the Laplace equation, equation 16.83 can be written as

$$\nabla \cdot V = -M \nabla^2(p + \rho gz) = 0 \quad (16.84)$$

where M is κ/η . This equation is characteristic of many diffusion problems that represent potential flows. Normally, these types of analogs are isotropic and in order to find a solution to the equations, the proper boundary conditions have to be specified.

Upon analyzing the displacement of epoxy by air, Feder (1989) finds that there is a relation between the number of air monomers (N) present in fingers within soil, and the distance of the monomer (r_i) from the point of injection. The relation he develops involves the radius of gyration (R_g) and total number of monomers containing air (N_o), such that $R_g = (N_o^{-1} \sum_i r_i^2)^{0.5}$. The results of Feder's investigation reveal that there is indeed a number-radius relation, with the resulting data from different fluid types and different times all falling on a simple curve. The number-radius relation is expressed as

$$N(r) = N_o \left(\frac{r}{R_g} \right)^D f \left(\frac{r}{R_g} \right) \quad (16.85)$$

where f is a crossover function that is assumed constant at $x < 1$, and approaches x^{-D} for $x > 1$. As a result, $N(r) \rightarrow N_o$ for $r \gg Rg$. Using this method, the results are best fitted with a fractal dimension D , which represents the fingering of the fluid within the media.

There are differences between the results obtained with a Hele–Shaw analog and those observed by viscous fingering. Since the Hele–Shaw has only a small width b (length scale), the fluid flow is controlled by this microscopic length. However, for a soil with a pore diameter equal to b , fluid flow is controlled in all directions by the microscopic length scale. Also, the boundary conditions chosen for each method have a significant influence on the results obtained. For the Hele–Shaw cell, the length plane is set by capillary forces, whereas in a soil the length scale is set by pore diameter. Thus, we can easily ascertain that for the Hele–Shaw analog flow, control is simply a matter of pressure distribution. For a soil, fluid flow or displacement is not simply a pressure difference, but the pressure relative to the capillary pressure at the pore neck. This is basically the same relation as that of the “inkbottle” effect (discussed earlier in the text), where it is more difficult for a fluid to enter a narrow neck versus a wide neck. Because soils are heterogeneous, pore necks vary in both size and shape throughout the system, which introduces a degree of randomness; without this randomness, a fractal structure cannot be produced. In both the Hele–Shaw analog and soil, flow of high viscosity fluids is controlled by the Laplace equation, thus fingering in soil has the components of the Laplace equation as it relates to Darcy’s law and pore geometry. The combination of the two components causes the generation of a fractal structure that can be analyzed by the method described by Feder (1989).

Earlier we discussed how the DLA process is one of random walking, where each particle eventually comes to rest and causes the formation of a fractal structure. We assume that the particles diffuse at a constant rate as described by the Einstein relation (equation 16.68). Using this concept, the random walk can be described by the basic diffusion equation

$$\frac{\partial C(r, t)}{\partial t} = D \nabla^2 C(r, t) \quad (16.86)$$

where $C(r, t)$ is the concentration of random walkers. By assuming steady state, $\partial C/\partial t = 0$ and equation 16.86 reduces to the Laplace equation. The velocity, as described by Witten and Sander (1983) is given by

$$V_{\perp} = -Dn \cdot \nabla C|_s \quad (16.87)$$

where V is perpendicular to the surface \perp , with surface normal n , and the equation is evaluated at the surface, $|_s$.

It is well known that if the displacing fluid has a lower viscosity than the resident fluid, wetting-front instability ensues and fingering will occur. For this case, the front moves with velocity $M\nabla(p + \rho gz)$. When the displacing fluid has a higher viscosity, the moving front is stable with a fractal dimension of 1. Also, the capillary number of a fluid greatly influences its ability to form fractal structures. The capillary number (Ca) is defined as

$$Ca = \frac{V\eta}{\sigma} \quad (16.88)$$

where V is the velocity, η the dynamic viscosity, and σ is the interfacial tension. The Ca is a measure of the ratio of capillary to viscous forces. From equation 16.88, it can be seen that Ca is increased by velocity (within a practical limit) or the use of fluids with a small interfacial tension. Thus, with high Ca , DLA accurately describes fingering in two-dimensional soil (i.e., $Ca = 0.05$ to about 0.1). For low Ca ($Ca \approx 10^{-4}$), the fractal structures formed are characteristic of those found using invasion percolation. This means, essentially, that the capillary forces completely dominate the viscous forces due to low velocity, and perhaps high viscosity. In invasion percolation (where Ca drastically decreases), pressure drops in the displacing,

resident fluids are neglected, and a simple pressure difference between the two fluids is calculated. This is accomplished by subtracting the pressure of the resident fluid (p_r) from the pressure of the displacing fluid (p_d) as follows

$$(p_d - p_r) = \frac{2\sigma \cos \phi}{r} \quad (16.89)$$

where σ is the interfacial tension between the two fluids, ϕ is the liquid-contact angle between the pore wall and the interface, and r is the pore radius. We note the similarity between equation 16.89 and the height of capillary rise. Perhaps the most prominent example of invasion percolation is the displacement of oil by water—two immiscible fluids.

SUMMARY

We have determined that fractals can provide meaningful answers to some of the questions now confronting water sciences, because they extend the principles and concepts used in geostatistics and scaling. This short discussion was meant to spark the interest of the reader. There are many possibilities to be explored for the use of fractals in the environmental sciences. However, the use of fractal mathematics in the earth sciences is still in its infancy.

ANSWERS TO QUESTIONS

- 16.1.** A normal probability distribution is the result of additive effects of numerous small, random, independent sources of variability. Thus, the properties for which we expect normal distribution are bulk density, water content at saturation, water content at -100 kPa, and the particle-size analysis.
- 16.2.** A log-normal distribution is the result of multiplicative effects due to spatial variability, thus the physical properties we expect to be log-normally distributed are scaling coefficient; saturated and unsaturated conductivity; diffusion coefficient; electrical conductivity; and pore-water velocity. However, in specific cases, the given property may not conform to the distribution given here. The answer, as to what makes a log-normal distribution more appropriate than a normal distribution for certain properties, is left to the student; please consult any standard statistics text.
- 16.3.** (a) Sampling error is a difference between a sample's average (resulting from the sampling process), and the population average (due to the heterogeneity of the property being studied). Measurement error is the difference between repeated results of a measuring process applied to a constant uniform object or property. (b) Statistical true value is the limiting mean of n measurements (that can be inaccurate due to bias) as n becomes large. Scientific true value is the actual value of an object of measurement by definition of the object. (c) scientific bias is a characteristic difference between the statistical true value and the scientific true value due to inadequate specification of the measurement process. Measurement bias is a characteristic tendency of results being either too high or too low, due to deficiencies in sampling apparatus or sampling materials. Sampling bias is the tendency of results to be in error due to inadequate sampling procedures or deficiencies in the sampling process.
- 16.4.** This could be done when the lower-precision method is less biased than the higher-precision method, and also when the lower-precision method is more efficient than the higher-precision method.

ADDITIONAL QUESTIONS

- 16.5.** Prove or show that the ruler dimension equals the similarity dimension in a triadic von Koch curve.
- 16.6.** What are the three basic characteristics of fractals?

- 16.7. How is the precision of an estimate of a derived value related to the precision of its component measurements?
- 16.8. What information is carried by the standard error? The coefficient of variation? When would the use of each be appropriate?
- 16.9. What information is displayed by a correlogram? A variogram?
- 16.10. Scaling is normally done on a computer; consequently, the method the computer uses to obtain a reduced SS is often misunderstood. Two intact soil cores are extracted and the following water retention data is obtained.

Core	h (cm)	S
1	10	0.95
1	20	0.90
2	10	0.94
2	20	0.93

Using the data from the table: **(a)** estimate the reference curve $h_m(s)$ by multiple-linear regression from *SAS* or similar statistical analysis; and **(b)** find coefficients a of the equation that best fit the data. This is accomplished by bringing to one side (set = 0)—that is, residual sums of squares—such that

$$\log(h_r) - a_0 - a_1 S \dots = 0$$

which is the ideal SS . (*Hint*: see section 16.2 and note equations 16.50–16.52.)