

UC Irvine

UC Irvine Previously Published Works

Title

Applications of the random model of drainage basin composition

Permalink

<https://escholarship.org/uc/item/6157833x>

Journal

Earth Surface Processes, 1(3)

ISSN

0360-1269

Authors

Smart, JS

Werner, C

Publication Date

1976-07-01

DOI

10.1002/esp.3290010304

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

APPLICATIONS OF THE RANDOM MODEL OF DRAINAGE BASIN COMPOSITION

J. S. SMART

IBM Thomas J. Watson Research Center, Yorktown Heights, New York, U.S.A.

AND

C. WERNER

School of Social Sciences, University of California, Irvine, California, U.S.A.

SUMMARY

The random model of drainage basin composition is founded on the assumptions that (a) natural channels are topologically random in the absence of geological controls and (b) for channel networks developed in similar environments, the exterior and interior link lengths are independent random variables with a common distribution for each type. The effectiveness of this model in estimating the values of geomorphic variables and in explaining and predicting geomorphic relationships is illustrated by several examples. The data required for these examples were obtained from map studies of 30 channel networks, comprising a total of about 8700 links, in eastern Kentucky. A common factor in the success of all three applications of the model is the way in which the planimetric features of drainage basins are determined by their underlying topologic structure.

KEY WORDS Channel networks Drainage basins Stochastic models

INTRODUCTION

The development and maintenance of drainage networks involve many complex processes, such as weathering, soil creep, fluid flow, and sediment transport. Although the physical bases of these processes are reasonably well understood, the overall complexity is such that an exact deterministic explanation of individual stream network properties appears impossible (Krumbein and Shreve (1970) p. 40, Shreve (1975)). It should however be possible to provide a good statistical description of the properties of drainage basins developed in a given environment. The observed values of some particular variable, such as drainage density, will vary randomly from basin to basin but will belong to a well-defined probability distribution determined by the local climate and geology. In addition, correlations between pairs of geomorphic variables are frequently relatively insensitive to variations in the environment and may persist in samples of data obtained from a rather wide range of climates and lithologies. As Shreve (1975) has noted, these points are clearly illustrated by the various empirical relationships between geomorphic variables that have been reported in the literature. Examples of such regularities are the linear relation between drainage density and the square of stream frequency (Melton (1958)) and the 0.6 power law dependence of mainstream length on area (Hack (1957)). Although such relationships are often stated in deterministic form, they are almost always obtained from regression analyses, and the results of individual measurements are seen to scatter randomly about the regression line. The deviations are consistently larger than is to be expected from measurement errors alone and can be attributed almost entirely to geomorphic randomness.

Such regularities provide useful information about drainage basin composition but they remain for the most part a series of isolated results, giving little hint of the complex interrelations of geomorphic

Received 1 March 1976

variables in drainage networks. A recent attempt to provide a more unified view of such phenomena is the random model of drainage basin composition which stems from a paper by Shreve (1966) and was developed by Shreve, Smart and others. A review of the earlier contributions is given by Smart (1972). This model rests on two basic postulates:

1. In the absence of geological controls, channel networks are topologically random (Shreve (1966) p. 27).
2. For drainage networks developed under similar environmental conditions, the exterior and interior link lengths are independent random variables with a single common distribution for each type (Smart (1968) p. 1005, Shreve (1967) p. 184, (1969) p. 402)).

These postulates are generally supported by observations on natural networks, although deviations from them have been noted for certain specific areas and in certain specific details. In general, they have been remarkably successful in providing quantitative and semi-quantitative explanations of a wide range of drainage network properties without the need of adjustable parameters (Werner and Smart (1973) p. 290–2, Smart (1973), Shreve (1975)).

A broad conceptual model such as this one may be expected (a) to produce better and more efficient procedures for estimating geomorphic properties, thus reducing the labour involved in measurement and computation; (b) to explain previously observed empirical relations; and (c) to predict new relations between geomorphic variables. Some specific examples of such applications of the random model appear in the references cited above. The purpose of this paper is to present a more detailed and comprehensive account of the effectiveness of the model for all three cases, using a large body of network data obtained from a single homogeneous region.

TOPOLOGIC PROPERTIES OF CHANNEL NETWORKS

The topologic properties of channel networks occupy an important place in the random model. In this section, the terminology for topologic properties, largely due to Shreve (1966), is reviewed. Some analogies between topologic and geometric variables are noted.

Figure 1 shows a hypothetical channel network. *Sources* are the points farthest upstream and the *outlet* is the point farthest downstream. A point at which two channels combine to form one is called

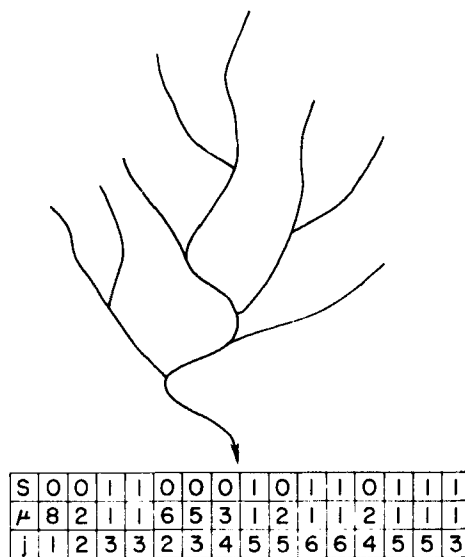


Figure 1. Example of a channel network with 8 sources. S is the topologic vector, μ the magnitude vector, and j the path length vector

a *junction*. It is assumed that no more than two channels combine at a single junction; apparent exceptions must be resolved by remapping or by an arbitrary, but usually unimportant, decision. *Links* are the segments of a channel network between a source and the first junction downstream, between two successive junctions, and between the outlet and the first junction upstream. In the first case, they are called *exterior* links and, in the last two cases, *interior* links.

A channel network with n sources has n exterior links, $n-1$ interior links and $n-1$ junctions. The *magnitude* of a link is the number of sources upstream; thus the magnitude of an exterior link is 1 and the magnitude of an interior link is the sum of the magnitudes of the two links that join at its upstream end. The magnitude of a channel network is that of its outlet link (also, the number of sources in the network).

Path length is another useful topologic property of channel networks. Each junction and source of a channel network can be characterized by its path length, or number of links connecting it to the outlet. Paths are classified as exterior or interior, depending on whether their upstream end is a source or a junction. Thus a channel network of magnitude n has n exterior paths and $n-1$ interior paths. The maximum path length, which must of course be associated with an exterior path, is called the *diameter*. For every network, there is an even number of exterior paths having the maximum length.

Jarvis (1972) first suggested that path length properties, particularly diameter and mean exterior path length, be used in classifying the topologic properties of channel networks, and a theoretical basis for the suggestion was later provided by Werner and Smart (1973). Perhaps the most convincing argument for using path lengths is that some of them can be identified as topologic analogues of important geometric channel lengths, e.g., diameter and mainstream length, exterior path length and flowpath length. This observation forms the basis of some of the estimation procedures described in a later section.

Finally, the topologic properties of any channel network can be completely specified by a string of ones and zeroes derived in the following way (Smart (1970)). Begin at the outlet and go around the network, turning left at every junction and reversing direction at every source. As each link is traversed for the first time (upstream direction), a one is written for an exterior link and a zero for an interior link; nothing is recorded for the downstream traversal. This method is essentially the same as those proposed by Scheidegger (1967) and Shreve (1967) but is more convenient for computer processing. As indicated in Figure 1, other link-associated properties, such as magnitude and path length, can be specified by strings of numbers (vectors) recorded in the same sequence.

COLLECTION AND HANDLING OF DATA

In discussing applications of the random model, it is advantageous to have a substantial set of network data obtained from a single region with homogeneous climate and geology. The region we selected for this purpose is in eastern Kentucky and was previously studied by Krumbein and Shreve (1970). They describe it as follows:

The topography of this area is mature, with steep slopes and narrow winding valleys and ridges . . . The drainage pattern is dendritic and shows no sign of lines of weakness such as joints or other geological controls . . . The bedrock consists of flat-lying (dips less than 50 feet per mile) relatively homogeneous Pennsylvanian sandstone and interbedded siltstone, shale, underclay, and coal. Poorly defined benches and somewhat broadcrested ridges probably attributable to structural control are present, but are widely scattered and nonpersistent. Thus the area, though not perfect, appears to be a good example of a mature landscape developed in the absence of geologic controls.

Thirty drainage basins were selected from the 1:24000, 40-ft contour interval USGS topographic maps of Offutt, Inez, Kermit, Lancer, Thomas, Varney, Williamson, Canoe, Haddix, and Noble quadrangles. The channel networks were outlined by the Strahler contour crenulation method. Link lengths were measured to the nearest 1/40 in (50 ft = 15.24 m full scale) with an architect's scale, curved

sections being approximated by a series of straight line segments. In practice, subdivision into segments was required for only about 10 per cent of the exterior links and an occasional interior link. For each basin, the total drainage area (A), total exterior link drainage area (A_e) and total interior link drainage area (A_i) were measured with a polar planimeter. The network magnitudes ranged from 43 to 267 and the basin areas from 1.84 to 13.74 km² (0.71 to 5.31 mi²). Table I gives descriptive statistics of some of the geomorphic properties that are expected to be relatively independent of magnitude and area. Here D is the drainage density (channel length per unit area), F the link frequency (number of links per unit area), K the Melton parameter ($= D^2/F$), λ the ratio of mean exterior link length to mean interior link length, and α the ratio of mean exterior link drainage area to mean interior link drainage area.

Both D and F , somewhat contrary to expectation, show small but statistically significant (0.05 level) decreases with increasing area. The hypothesis of a negative correlation between drainage density and area is further supported by the fact that the mean value of D , 5.80 km⁻¹, is about 20 per cent less than the value of 7.3 km⁻¹ reported by Krumbein and Shreve (1970) for 10th magnitude basins in the same region. A decrease in drainage density implies an increase in mean link length, which of course violates the second postulate of the random model. However, the purpose of this paper is not to carry out further tests of the random model, but rather to see if it is a sufficiently good approximation to natural networks to be used effectively in the three applications listed in the introductory section.

The three dimensionless variables show no significant dependence on area. The link length ratios are quite comparable with those that can be inferred from results of previous investigations. For example, data from 46 networks in the western United States (Melton (1957)) gives a mean of 1.54 and a standard deviation of 0.30, while data from 60 networks in southern Indiana (Coates (1958)), in a region which is geologically rather similar to this one, gives corresponding values of 1.46 and 0.40. The previously-reported data on α is so scant that no meaningful comparisons can be drawn. One might expect that λ and α would be positively correlated and, indeed, the Spearman rank correlation coefficient is 0.631, which value is significant at the 0.01 level.

Table I. Geomorphic properties of 30 drainage basins

	$D(\text{km}^{-1})$	$F(\text{km}^{-2})$	K	λ	α
Mean	5.80	44.1	0.773	1.43	1.34
Standard deviation	0.78	10.3	0.090	0.21	0.28
Coef. of variation	0.135	0.233	0.116	0.148	0.208
Minimum	4.74	28.6	0.599	1.11	0.80
Maximum	7.97	70.1	0.929	1.97	1.98

The 30 channel networks contain 4377 exterior links and 4347 interior links. Some statistical properties are given in Table II and the frequency distributions are shown in Figure 2. As required by the second postulate of the random model, the exterior and interior link lengths appear to be drawn from different populations.

Table II. Link length statistics

	Exterior	Interior
Sample size	4377	4347
Mean (m)	158.6	111.5
Standard deviation (m)	84.6	81.3
Coef. of variation	0.533	0.729
Minimum (m)	30	0
Maximum (m)	945	745

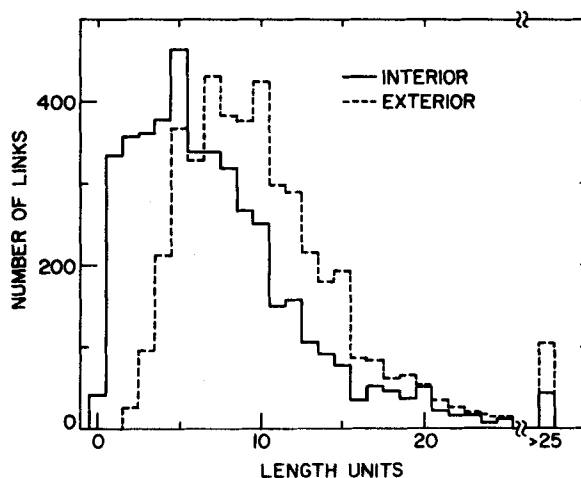


Figure 2. Frequency distributions for exterior and interior path lengths. The unit of length is 15.24 m

Accuracy and reproducibility of map data in studies such as this one are of course of paramount importance. Ideally, it is desirable that each source located on the map correspond to a natural source and that no natural sources be overlooked. This ideal can be approximated for some maps in some areas but as Krumbein and Shreve ((1970) p. 12-13) point out, it is usually necessary to accept somewhat lower standards. Useful and consistent information can still be obtained if the sources and channels are identified by reasonably objective map criteria and can be reproduced to reasonable accuracy by different operators or by the same operator repeating an earlier study. Krumbein and Shreve ((1970) p. 19-38) have made a detailed investigation of operator variation in map measurement for this same region in eastern Kentucky. We undertook a similar but much less extensive study, with results that were substantially the same as those of Krumbein and Shreve. As the network is extended upstream from the blue lines, the most crucial decision is whether a given set of crenulations in successive contour lines should be identified as a channel. We found that the operator variation in determining the number of exterior links was almost always less than 5 per cent. This is important, because if two operators identify essentially the same set of exterior links, then their results for the network topology and for the interior link lengths will also be essentially the same. We were less successful in obtaining consistent identification of the sources, or upstream ends of exterior links, and the length data for exterior links should be regarded as less reliable than that for interior lengths. All of the data used in this paper were collected by a single person, Thomas McElroy, a student at the Department of Geological Sciences, SUNY, Binghamton, New York.

For each of the 30 networks, we constructed the topologic binary vector, as described in the previous section, and a link length vector, in which the individual link lengths (in units of 1/40 in) were entered in the order in which the corresponding links appear in the binary vector. These data were stored in the APL SV system of the 360/91 computer at the IBM Thomas J. Watson Research Center. Other topologic and channel length information of interest can then be easily obtained by various simple algorithms for manipulating the binary and length vectors (Smart (1970)). Among the results derived in this way were the diameters (d), total exterior path lengths (p_e), mean exterior path lengths (p_e/n), mainstream lengths (L'), total flowpath lengths (P_E), total channel lengths (L), mean exterior link lengths ($\bar{l}_e$), and mean interior link lengths ($\bar{l}_i$) for each network. With the inclusion of the area measurements, such properties as drainage density and link frequency were also computed. Similar procedures for storing and analyzing data on channel networks were developed independently by Jarvis (1975) in a study of drainage basins in two regions of Great Britain. Copies of our data are available upon request to the senior author.

ESTIMATION

Channel Lengths

In searching for relations between hydrograph parameters and the physical characteristics of watersheds, hydrologists have frequently used various functions of the flowpath lengths (distances from outlet to sources measured along the channel) including the sum, mean, standard deviation, coefficient of variation, and maximum (e.g., Rice (1970), Rogers (1970)). The maximum flowpath length is of course more commonly called the mainstream length. It seems intuitively apparent that flowpath lengths and hydrograph properties should be rather closely related, and the limited number of such investigations is no doubt in part due to the considerable amount of labour required in making the necessary measurements even for watersheds of moderate size. Various approximate methods which involve less effort have been proposed; see for example, Busby and Benson (1960) and references cited there.

Werner and Smart (1973) and Shreve (1975) have recently made use of an approximation procedure in which the geometric channel lengths are replaced by their topologic analogues. For example, diameter is substituted for the mainstream length and exterior path length for flowpath length. This estimation method permits a drastic reduction in labour since determination of topologic lengths, either by direct counting or by use of the binary vector, is much quicker than making the corresponding geometric length measurements. The accuracy of estimation compares favourably with previous approximation methods and, unlike the other methods, it can be determined by application of standard statistical theory.

The general idea is that the length of a channel segment containing m links can be satisfactorily estimated by the relation

$$L(m) \approx m\mu \quad (1)$$

where μ is the expected value of the population of link lengths for that environment. Since the m links can be regarded as a random sample of the population (second postulate), application of the central limit theorem of statistics (Hoel (1962) p. 145) yields the result that, for sufficiently large m , the error of estimation $L(m) - m\mu$ is normally distributed with a mean of zero and a variance of $m\sigma^2$ where σ^2 is the population variance. Thus the accuracy of the approximation depends, as might be expected, on the number of links and the length distribution. Shreve ((1975) p. 1169) gives the explicit result

$$m = (zC_v/k)^2 \quad (2)$$

where z is a standard normal variate corresponding to a given confidence level, k is a fractional error in the estimate, and C_v is the coefficient of variation of link length distribution. As a specific example, if we want the error to be within ± 10 per cent ($k = 0.1$) at least 95 per cent of the time ($z = 1.96$), we must have

$$m = 384 C_v^2 \quad (2a)$$

or greater.

For networks of the magnitude studied here ($n > 40$), both the mainstream length and flowpath length have many more interior than exterior links, and only negligible error will be introduced if we assume they are constituted entirely of interior links. If we further assume that the sample of 4347 interior link lengths adequately represents the population distribution for the eastern Kentucky region, we find by substituting $C_v = 0.729$ from Table III that about 200 links are required to obtain the specified accuracy.

If the entire body of data obtained for hydrologic or geomorphic investigations comes from networks developed under the same environmental conditions (as in this paper), the population mean interior link length is the same for all basins. Thus all approximations of the form of equation (1) contain the same constant factor μ . For many kinds of statistical analysis such constant factors are unimportant, and channel lengths can be replaced by their corresponding topologic variables, thus eliminating the necessity for any length measurements at all. Table III gives the correlation coefficients between L' ,

P_E , flowpath standard deviation S_E and their respective topologic analogs, d , p_e , and exterior path length standard deviation s_e . The correlation coefficients are all greater than 0.9. Whether this level of correlation is satisfactory depends on the details of the investigation, but a survey of previous hydrologic and geomorphic studies suggests that it would be readily acceptable in the large majority of cases. It should be noted that in the comparison of L' and d , the geometric and topologic measurements can refer to different exterior paths. In practice, however, this situation occurs less than 10 per cent of the time, and in most of the cases the difference produced is small.

If the data represent different regions or if an absolute estimate of the geomorphic variable is desired, then appropriate values of the mean link length μ must be introduced. Table III shows the results of fitting the geometric variables L' , P_E , and S_E to an equation of the form

$$Y = a + b \mu X \quad (3)$$

where Y is the geometric variable, X the corresponding topologic variable, and $\mu = 111.5$ m from Table II.

Table III. Relations between geometric and topologic variables

Geometric Var. (Y)	Topologic Var. (X)	Metric Factor (g)	Corr. Coef.	Regression parameters		
				a	$a/\bar{Y}$	b
L'	d	$\bar{l}_i$	0.903	0.287	0.064	0.950
P_E	p_e	$\bar{l}_i$	0.981	11.3	0.027	0.994
S_E	s_e	$\bar{l}_i$	0.924	0.035	0.033	0.955
β_G	β_T	$\bar{l}_i^2/\bar{a}_0$	0.795	0.593	0.102	0.851
D	$F^{1/2*}$	$K^{1/2}$	0.902	-0.431	-0.074	1.069

* Not a topologic variable.

(In the more general case in which the data are drawn from different regions, μ will of course vary from one observation to another.) For unbiased estimation, $a = 0$, $b = 1$, and we see that this condition is satisfactorily approximated; the value $a/\bar{Y}$ is listed as an appropriate scale-free indicator of how near a approaches zero. Figure 3 shows plot of Y versus its estimator μX for all three cases.

One practical limitation of this method of approximation is the relative accuracy of the value used for μ . Most workers will not wish to measure 4347 link lengths and will be content to estimate the expected value from some smaller sample, say of size j . The relation corresponding to equation (1) can then be written

$$L(m) = m\bar{X}_m \approx m\bar{X}_j \quad (1a)$$

where $\bar{X}_k$ is the mean of a random sample of k links. The central limit theorem can again be used to show that the error $m(\bar{X}_m - \bar{X}_j)$ is normally distributed with a mean of zero and a variance $m(1 + m/j)\sigma^2$. Thus by comparison with the previous results, we see that the use of a sample of size j to estimate μ produces no bias but increases the variance by a factor of m/j . This analysis is of rather limited worth, however, since it is rather unlikely that σ^2 will be known when μ is not. An alternative procedure which avoids this objection is to use the statistic

$$T = (\bar{X}_m - \bar{X}_j)[mj(m + j - 2)/(m + j)(ms_m^2 + js_j^2)]^{1/2} \quad (4)$$

where s_m and s_j are the sample standard deviations. T is distributed as Student's t with $m + j - 2$ degrees-of-freedom. Details are given in Hoel ((1962) p. 276-7).

Two further cautionary notes should be added. In some cases of interest, it may be that m is not 'large' so that use of the central limit theorem to determine error distributions is unjustified. For example in our study, the diameters ranged from 18 to 62. The recourse here is again to use the statistic T defined above, since the theory behind it does not require any assumptions about sample size. Finally,

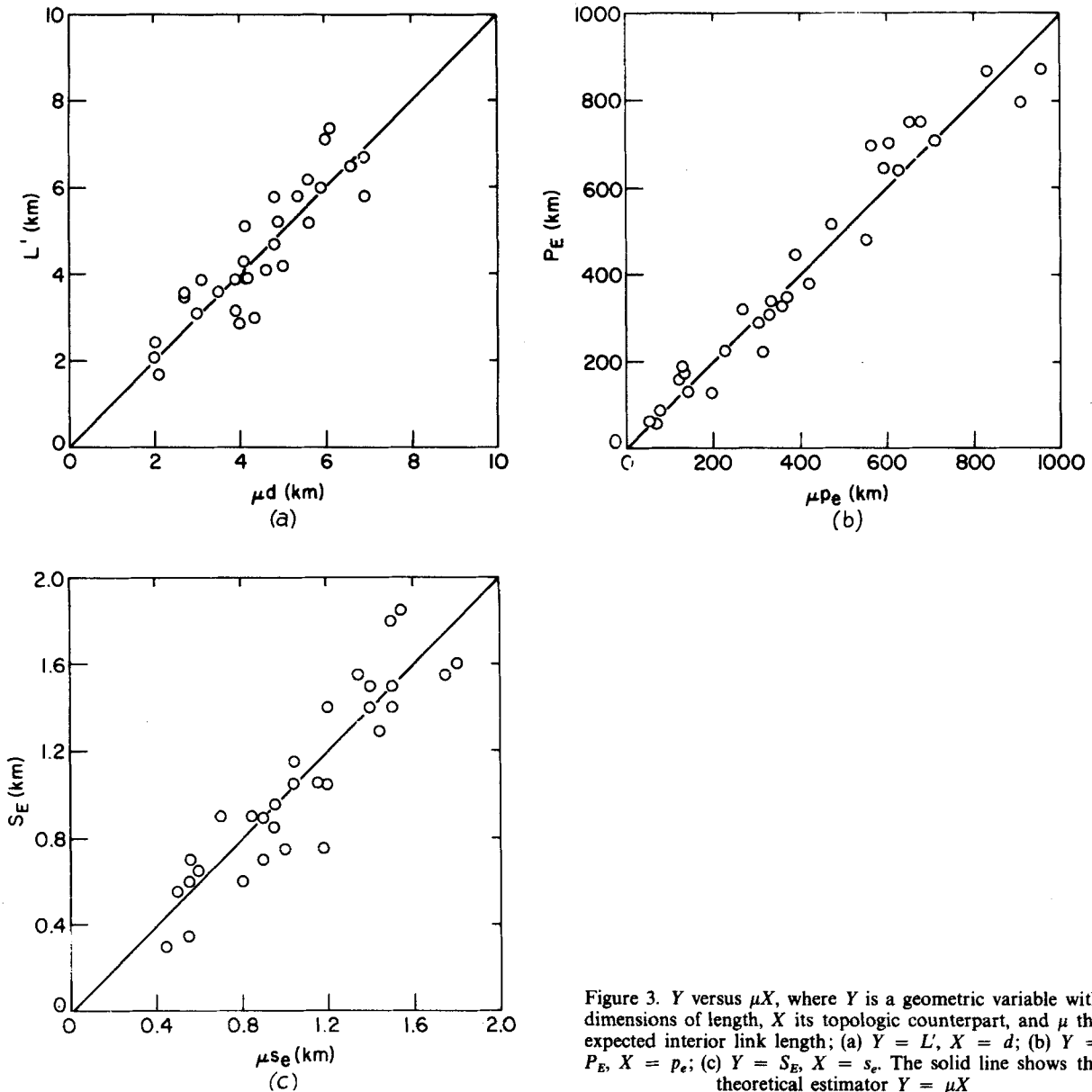


Figure 3. Y versus μX , where Y is a geometric variable with dimensions of length, X its topologic counterpart, and μ the expected interior link length; (a) $Y = L'$, $X = d$; (b) $Y = P_E$, $X = p_e$; (c) $Y = S_E$, $X = s_e$. The solid line shows the theoretical estimator $Y = \mu X$.

it should be noted that the value of P_E contains multiple contributions from each link according to its magnitude, and that the total number of independent link lengths involved is $2n - 1$.

Shape factors

This estimation procedure is of course not limited to channel lengths. It is only necessary to replace equation (1) by the more general relation

$$Y \approx gX \quad (1b)$$

where X and Y are corresponding topologic and geometric variables, respectively, and g is an appropriate metric factor, whose value is normally determined by sampling. The error analysis is carried out by

an obvious extension of the method described for channel lengths, but since the results are highly dependent on the individual case, we shall not discuss it further here.

As an example, we consider the problem of estimating the dimensionless quantity

$$\beta_G = (L')^2/A \quad (5a)$$

This variable, or some quantity proportional to it, has frequently been used as a watershed variable in hydrologic analysis (e.g. Benson (1962)). It is a measure of drainage basin shape; if two watersheds of the same area have different β_G , the larger value implies a more elongated form and consequently a greater time of concentration and a lower peak discharge.

The topologic analog of β_G can be obtained by replacing L' by $d\bar{l}_i$ and A by $(2n-1)\bar{a}_0$ where $\bar{a}_0$ is the mean link drainage area including both interior and exterior links. Thus

$$\beta_T = d^2/(2n-1) \quad (5b)$$

and the metric factor g is $\bar{l}_i^2/\bar{a}_0$. Note that although β_G is dimensionless, its estimator is not independent of the geometric properties of the network. For our data, $g = 0.520$. The least squares fit to the equation

$$\beta_G = a + bg\beta_T \quad (6)$$

yields $\rho = 0.795$, $a = 0.59$, $a/\bar{\beta}_G = 0.10$ and $b = 0.851$. Thus both the scatter and bias are worse than in the cases previously discussed but perhaps not so bad as to preclude a cautious use of the estimator. Moreover, it may well be that β_T itself is as good or better than β_G in characterizing shape and shape-related properties; this possibility can be resolved only by appropriate tests.

Many other shape factors in addition to β_G have been proposed for use in hydrologic and geomorphic analysis. One closely related group involves various functions of L_b/A^2 , where L_b is the basin length. These include the Horton (1932) $F_H = L_b/A^2$, the Corps of Engineers (1949) S , the Schumm (1956) elongation ratio E , and the Chorley-Malm-Pogorelski (1957) lemniscate factor k . As Jarvis (1975) has pointed out, these four factors are all equivalent except for minor differences in the definition of L_b . If such differences are ignored, we have

$$F_H = 1/S = \pi E^2/4 = \pi/4k \quad (7)$$

Since d should be a good measure of basin length as well as mainstream length, it seems likely that β_T would be an acceptable substitute for any of the above factors.

Drainage density

Drainage density is an important variable in almost any hydrologic or geomorphic study. Although accurate measurements are required in some investigations, in many others a good approximation would be quite satisfactory. Most of the labour of measurement is involved in determining the total channel length, L .

Estimates of sufficient accuracy can often be obtained from the relation

$$D = ((2n-1)K/A)^{1/2} \quad (8)$$

which is derived by combining the definitions of D , F and K . We assume that n and A are known (their values must be determined in any event) and use an approximation for $K^{1/2}$ obtained from sampling. From Table I we see that K is a highly conservative quantity; $K^{1/2}$ has a coefficient of variation of 0.058 so that only a few sampling measurements are required to obtain an accurate approximation. By using the mean value of $K^{1/2}$ for our 30 networks, we find that the correlation coefficient between measured and estimated values of D is 0.902 and that the coefficients in the corresponding regression line are $a/\bar{D} = -0.074$ and $b = 1.069$. Thus the level of estimation is quantitatively very similar to that illustrated in Figure 3 for the length variables. Specific calculations show that 28 of the 30 estimates are within ± 10 per cent of the measured value and 18 are within ± 5 per cent of the measured value.

Note that this procedure differs from those discussed previously in that the estimator $F^{1/2}$ is not a pure topologic variable. Moreover it may seem somewhat illogical to use $K^{1/2} = \bar{l}_0/\bar{a}_0^{1/2}$ as the

metric factor and $F^{1/2} = 1/\bar{a}_0^{1/2}$ as the estimator. This choice is, however, dictated by the fact that $K^{1/2}$ is more tedious to measure than $F^{1/2}$ but at the same time has an appreciably lower coefficient of variation.

The recommendation for this procedure must be qualified with the observation that rather little is known about the general properties of the variable K . Melton (1960) and Smart (1973) have noted that its value is almost always rather close to unity, and Smart (unpublished results) in studying several different regions with dendritic drainage networks found means ranging between 0.70 and 1.14 and coefficients of variation ranging between 0.05 and 0.12. It should also be noted that this method is probably best suited for estimating individual variations in drainage density among basins in a single homogeneous region. For comparing drainage densities in different regions, which would require multiple determinations of K , the line intersection method proposed by Carlston and Langbein (unpublished work) and discussed by McCoy (1971) and Mark (1974) would appear to involve less labour and be of comparable accuracy.

EXPLANATION

The distinction between explanation and prediction is in one sense not very profound, since it mainly rests on the order in which observations and theoretical analyses were carried out. We feel however that the point of view in the two procedures is sufficiently different to warrant separate discussions. We first consider some uses of the model in explaining previously observed relations between pairs of geomorphic variables.

Most of the earlier papers on the random model devoted some space to explaining Horton's laws of drainage composition and related phenomena. This work has been reviewed previously (Smart (1972)) and will not be discussed further here. Also, as previously noted (Smart (1973), Werner and Smart (1973)), the explanation and prediction of relations between Horton parameters is somewhat unrewarding (except insofar as it provides a test of the model) because their behaviour is frequently determined more by the ordering scheme than by fundamental network properties.

The set of geomorphic variables used to characterize a drainage basin are known to be rather highly intercorrelated, e.g. Doornkamp and King ((1971) Part 1). In many cases, the quantitative relation between a particular pair (Y_1, Y_2) can be rather well represented, at least over some limited range, by a power law expression

$$Y_1 = aY_2^b \quad (9a)$$

Values of a and b are usually estimated from the data by a regression analysis of the logarithms

$$\ln Y_1 = \ln a + b \ln Y_2 \quad (9b)$$

If Y_1 and Y_2 are planimetric variables, it frequently happens that their interdependence stems from an underlying relationship between the corresponding topologic variables X_1 and X_2 .

The general approach in this section is to use the second postulate of the random model, as before, to approximate Y_1 and Y_2 by g_1X_1 and g_2X_2 , respectively, where g_i is a metric factor. Then the first postulate (topological randomness) is employed to show that the relation between X_1 and X_2 can be satisfactorily expressed by

$$X_1 = a'X_2^{b'} \quad (10)$$

and that the correlation between X_1 and X_2 is approximately equal to that between Y_1 and Y_2 . As only a limited number of results for the random model can be obtained by strictly analytical procedures, Monte Carlo methods may be required to carry out these steps. Finally, we deduce approximate values of a and b by substituting g_iX_i for Y_i in equation (9a).

$$a \approx g_1a'/g_2^b \quad (11a)$$

$$b \approx b' \quad (11b)$$

These values are then compared with the regression results of equation (9b).

The power law relation between mainstream length and drainage area (Hack (1957)) provides a good illustration of the application of these procedures (Shreve (1970) (1974), Werner and Smart (1973)). Hack found that his measurements on 90 drainage basins in Virginia and Maryland and the results of Langbein *et al* (1947) on about 400 drainage basins in the northeastern United States could be represented by

$$L' = 1.5 A^{0.6} \quad (12)$$

when L' is expressed in km and A in km^2 . The correlation coefficient was about 0.9. Similar results were obtained by Gray (1961).

The topologic variables corresponding to L' and A are, respectively, the diameter d and the total number of links $2n - 1$, and the corresponding metric factors are $\bar{l}_i$ and $\bar{a}_0$. Shreve ((1974) p. 1169–71) gave a theoretical analysis of the variation of expected diameter with magnitude for topologically random networks. He showed that the slope of the $\ln E[d] - \ln n$ curve is unity for small n and almost certainly approaches $1/2$ as n becomes very large. In a range of intermediate magnitudes (10–500) which covers much of the literature on channel networks, the slope varies from 0.69 to 0.53. In order to make a detailed comparison, however, it was necessary to use a Monte Carlo method in which networks were drawn from a topologically random population and link lengths and link drainage areas were randomly assigned from a table of measured values. The agreement between simulated and measured values with respect to both trend and dispersion was excellent for basins of area less than about 10^4 km^2 . See Shreve's paper for a discussion and resolution of some discrepancies for larger basins.

Werner and Smart ((1973) p. 290–1) obtained results supporting Shreve's in a somewhat different way. They generated a sample of topologically random networks with magnitudes between 20 and 200 (which range falls in the middle of Hack's observations) and obtained the relation

$$d = 1.4(2n - 1)^{0.58} \quad (13)$$

from a regression analysis. This result was compared with equation (12) by using values of $\bar{l}_i/\bar{a}_0^b$ appropriate for Hack's study. The results ($a = 1.57$, $b = 0.58$) were again in excellent agreement with observation.

In the present paper, as in the works cited above, it was necessary to resort to simulation procedures to obtain theoretical results for a topologically random population. Since most of the relations of interest were magnitude-dependent, the magnitudes of the networks in our simulated sample were randomly chosen from a uniform distribution of integers between 40 and 270, thus matching the range of observations. Five hundred choices were made and for each choice one network was selected from the appropriate topologically random population for that magnitude.

Table IV gives the power law coefficients a and b and linear correlation coefficients ρ for several sets of observed geometric variables, observed topologic variables, and simulated topologic variables. The values of a listed for topologic variables were obtained from the a' values *via* equation (11a) and thus are quantitatively comparable. For all of the relations considered here, b and b' are expected to be equal, so that no transformation is necessary. The first line of the table merely confirms the previous extensive investigations (described above) of the L' - A relation. The other relations have not, to our knowledge, been studied before. Langbein *et al* (1947) defined a quantity rather closely related to P_E and found that it varied with the 1.56 power of A ; thus their result agrees more closely with the simulation value 1.5 than do our observations on P_E .

The one notable, and unexplained, discrepancy in the table is in fact in the dependence of P_E on A , where the exponent of 1.3 compares very poorly with the value near $3/2$ obtained for both the observed and simulated topologic variables. Otherwise, the agreement is quite good. In all cases, the regression coefficients for the simulation fall within the 95 per cent confidence limits of the corresponding results for the observed variables, both geometric and topologic, and in about half the cases they fall within the 50 per cent confidence limits. Also, the trends in correlation coefficients for the three kinds of data are generally the same. These results all support the hypothesis that the observed statistical relations between the pairs of geometric variables are largely consequences of the random model. As stated previously, two distinct effects are involved. The first postulate of the model determines the

Table IV. Power law relations between pairs of geometric and topologic variables

Variables				Coefficients								
Geometric		Topologic		<i>a</i>			<i>b</i>			ρ		
<i>Y</i> ₁	<i>Y</i> ₂	<i>X</i> ₁	<i>X</i> ₂	G*	TO	TS	G	TO	TS	G	TO	TS
<i>L'</i>	<i>A</i>	<i>d</i>	$2n - 1$	1.55	1.37	1.45	0.56	0.59	0.57	0.890	0.881	0.809
<i>P_E</i>	<i>A</i>	<i>p_e</i>	$2n - 1$	31.3	12.7	13.2	1.31	1.54	1.50	0.957	0.957	0.966
<i>S_E</i>	<i>A</i>	<i>s_e</i>	$2n - 1$	0.33	0.30	0.36	0.60	0.63	0.56	0.836	0.807	0.702
<i>P_E</i>	<i>L'</i>	<i>p_e</i>	<i>d</i>	18.2	16.8	18.6	2.00	2.13	2.01	0.922	0.952	0.917
<i>S_E</i>	<i>L'</i>	<i>s_e</i>	<i>d</i>	0.19	0.21	0.18	1.13	1.10	1.08	0.983	0.979	0.969
<i>S_E</i>	<i>P_E</i>	<i>s_e</i>	<i>p_e</i>	0.063	0.076	0.086	0.47	0.45	0.43	0.899	0.895	0.838

* G, TO, TS indicate results for geometric observations, topologic observations, and topologic simulations, respectively.

nature of the statistical relations between pairs of topologic variables and the second postulate transfers these relations, with relatively little modification, to the corresponding pairs of geometric variables.

PREDICTION

To illustrate the application of the random model in predicting geomorphological phenomena, the procedures of the previous section are essentially reversed. First, a question is posed about the properties of some particular geomorphic variable. Then the random model is used to predict these properties and, finally, observations on natural networks are used to test the prediction.

One unforeseen difficulty that arose in carrying out this programme was that of finding meaningful geomorphic variables which have not been previously investigated. It appears that quantitative fluvial geomorphologists have been both ingenious and diligent in their efforts to characterize drainage basins. Consequently, our examples of predictive applications of the random model are limited to two.

As the first example, we consider the question: what fraction of the total basin area is drained by the largest tributary to the mainstream? A limited poll of geomorphologists produced estimates of the mean ranging from 0.1 to 0.55. Thus this variable, though presumably of minimal theoretical or practical interest, is eminently suitable for our purposes. Not only are its properties uninvestigated, but they are also not immediately obvious even to persons with a comprehensive knowledge of drainage basins.

To make the question quantitative, consider the set of sub-basins draining into the mainstream and let the largest area be A_i . Similarly, let n_i be the largest magnitude in the set of subnetworks that join the diameter. Then the fractions of interest are

$$f_G = A_i/A \quad (14a)$$

$$f_T = (2n_i - 1)/(2n - 1) \quad (14b)$$

As in the case of mainstream length and diameter, it is possible for f_G and f_T to refer to different parts of the basin but, also as before, the overall effect is not very important.

Table V gives some descriptive statistics for f_T , both simulated and observed, and for f_G . The random model prediction appears to be very good in all respects, especially considering that the simulation results would be expected to have a greater range because of the larger sample size. Comparison of the distributions by the Kruskal-Wallis test yields a statistic of 0.34 while the corresponding χ^2 -value for 5 per cent confidence and 2 degrees-of-freedom is 5.99. Thus there is no reason to reject the hypothesis that the three sets of variates have the same distribution function.

As the second example, we use the random model to predict another suitably obscure phenomenon, the relationship between tributary area ratios and basin shape factors. The theoretical prediction, obtained by regression of f_T on β_T using the simulation data, is

$$f_T = 0.39 - 0.025 \beta_T \quad (15a)$$

Table V. Descriptive statistics of area ratios

	Topologic		Geometric
	Simulated	Observed	Observed
Sample size	500	30	30
Mean	0.263	0.245	0.243
Standard deviation	0.110	0.083	0.083
Coef. variation	0.419	0.340	0.341
Maximum	0.644	0.401	0.400
Minimum	0.090	0.070	0.119

with $\rho = -0.506$. The corresponding regression of observed values of f_T and β_T yields

$$f_T = 0.41 - 0.028 \beta_T \quad (15b)$$

with $\rho = -0.611$. The predicted intercept and slope of the regression of the corresponding geometric factors are, respectively, 0.39 and $-0.025\bar{a}_0/\bar{l}_i^2 = -0.048$. The actual result

$$f_G = 0.39 - 0.046 \beta_G, \quad (15c)$$

is closer than might be reasonably expected, considering the sample size and natural scatter in the data. The correlation coefficient is -0.570 .

Thus the random model gives an excellent prediction of both the trend and dispersion in the relations between the observed variables. Although the correlation coefficient for the simulated data is slightly smaller in magnitude than those for the observed data, the difference is quite comparable to that expected from finite sampling effects; Monte Carlo studies show that the standard deviation of estimates of ρ for samples of size 30 is about 0.09.

CONCLUSIONS

In this paper, we have presented a limited number of examples of the application of the random model in estimating geomorphic variables and in explaining and predicting geomorphic relationships associated with channel networks. Other examples of equal quality are readily available and have been omitted only in order to keep the manuscript to a reasonable length. We feel that these results provide convincing evidence of the usefulness of the random model in interpreting and understanding the planimetric properties of drainage networks developed in a homogeneous environment.

A geomorphologist or hydrologist may often be required to work with data selected from several different environments or from a single environment with strong geological controls, in which case the networks may not be topologically random. The question then arises as to whether the results and procedures described above are valid under such conditions. Because of the essentially infinite number of ways in which data can be mixed and in which networks can deviate from topological randomness, no single answer can be given to this question. A few remarks may help to put matters in perspective. The weight of evidence from this and other papers indicates that the applicability of the random model is not seriously affected by modest deviations from homogeneous data or from topological randomness. For example, the 0.6 power dependence of mainstream length on area could be detected in a set of data taken from a number of different regions because both the proportionality factor a and the exponent b (particularly the exponent) are only weakly dependent on geometric scale factors which vary from one environment to another. As another example, the unequal abundance of *cis* and *trans* links, which was discovered by James and Krumbein (1969) and seems to be a general feature of natural stream networks, points to a basic departure from topological randomness; it does not however have any important consequences for the results of this paper.

We also note that the methods used for estimation depend only on the second assumption of the random model, and thus may be more generally applicable than those used for explanation and prediction, which depend on both assumptions. This point is illustrated by another set of data which we

took on stream networks in the San Dimas Experimental Forest of the San Gabriel Mountains in southern California. The networks were clearly not drawn from a topologically random population. Instead the observed values of d and p_e tend to be too large, a kind of deviation consistent with that previously reported by Smart (1969) for networks in the western United States. A test of the estimation procedures for this region, however, indicates that the level of accuracy is as good as that for eastern Kentucky.

A central theme which appears many times in this paper is the way in which the planimetric properties of stream networks are determined by the underlying topologic structure. The same point has been noted by Shreve (1975), and we believe the results obtained here generally support Shreve's remarks and tend to refute the claims of Werrity ((1972) p. 193–4) and others that an understanding of topologic properties is of little practical value to the geomorphologist.

ACKNOWLEDGEMENT

We are pleased to acknowledge the valuable efforts of Thomas McElroy in collecting, coding, and processing the large quantities of data required for this study.

REFERENCES

- Benson, M. A. (1962). 'Factors influencing the occurrence of floods in a humid region', *U.S. Geological Survey Professional Paper* 1580-B.
- Busby, M. W., and Benson, M. A. (1960). 'Grid method of determining mean flow-distance in a drainage basin', *International Association of Scientific Hydrology Bulletin*, **20**, 32–36.
- Chorley, R. J., Malm, D. E. G., and Pogorelski, H. A. (1957). 'A new standard for estimating drainage basin shapes', *American Journal of Science*, **255**, 138–141.
- Coates, D. R. (1958). 'Quantitative geomorphology of small drainage basins in southern Indiana', *Project NR 389-042 Technical Report No. 10*, Department of Geology, Columbia University.
- Corps of Engineers, Department of the Army, Washington District 1949 Unit hydrograph compilations.
- Doornkamp, J. C., and King, C. A. M. (1971). *Numerical Analysis in Geomorphology*, London, Edward Arnold.
- Gray, D. M. (1961). 'Interrelationship of watershed characteristics', *Journal of Geophysical Research*, **66**, 1215–1223.
- Hack, J. T. (1957). 'Studies of longitudinal stream profiles in Virginia and Maryland', *U.S. Geological Survey Professional Paper* 294-B, 45–97.
- Hoel, P. G. (1962). *Introduction to Mathematical Statistics*, New York, Wiley.
- Horton, R. E. (1932). 'Drainage basin characteristics', *Transactions of the American Geophysical Union*, **13**, 350–361.
- James, W. R., and Krumbein, W. C. (1969). 'Frequency distributions of stream link lengths', *Journal of Geology*, **77**, 544–565.
- Jarvis, R. S. (1972). 'New measure of the topologic structure of dendritic drainage networks', *Water Resources Research*, **8**, 1265–1271.
- Jarvis, R. S. (1975). 'Law and order in stream networks', Ph.D. Dissertation, Cambridge University.
- Krumbein, W. C., and Shreve, R. L. (1970). 'Some statistical properties of dendritic channel networks', *Technical Report No. 13, ONR Task No. 389-150*, Department of Geological Sciences, Northwestern University; *Special Project Report, NSF Grant 6A-1137*, Department of Geology, University of California, Los Angeles.
- Langbein, W. B. *et al* (1947). 'Topographic characteristics of drainage basins', *U.S. Geological Survey Water-Supply Paper* 947-C, 125–155.
- Leopold, L. B., Wolman, M. G., and Miller, J. P. (1964). *Fluvial Processes in Geomorphology*. San Francisco, W. H. Freeman.
- Mark, D. M. (1974). 'Line intersection method for estimating drainage density', *Geology*, **2**, 235–236.
- McCoy, R. M. (1971). 'Rapid measurement of drainage density', *Bulletin of the Geological Society of America*, **82**, 757–762.
- Melton, M. A. (1957). 'An analysis of the relations among elements of climate, surface properties, and geomorphology', *Project NR 389-042 Technical Report No. 11*, Department of Geology, Columbia University.
- Melton, M. A. (1958). 'Geometric properties of mature drainage basins and their representation in an E_4 phase space', *Journal of Geology*, **66**, 35–56.
- Melton, M. A. (1960). 'Concept of state of a drainage system', *Bulletin of the Geological Society of America*, **71**, 1928.
- Rice, R. (1970). 'Storm runoff from chaparral watersheds', Ph.D. Dissertation, Colorado State University.
- Rogers, W. F. (1972). 'New concept in hydrograph analysis', *Water Resources Research*, **8**, 973–981.
- Scheidegger, A. E. (1967). 'On the topology of river nets', *Water Resources Research*, **3**, 103–106.
- Schumm, S. A. (1956). 'Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey', *Bulletin of the Geological Society of America*, **67**, 597–646.
- Shreve, R. L. (1966). 'Statistical law of stream numbers', *Journal of Geology*, **74**, 17–37.
- Shreve, R. L. (1967). 'Infinite topologically random channel networks', *Journal of Geology*, **75**, 179–186.
- Shreve, R. L. (1969). 'Stream lengths and basin areas in topologically random channel networks', *Journal of Geology*, **77**, 397–414.
- Shreve, R. L. (1974). 'Variation of mainstream length with basin area in river networks', *Water Resources Research*, **10**, 1167–1177.
- Shreve, R. L. (1975). 'The probabilistic-topologic approach to drainage-basin geomorphology', *Geology*, **3**, 527–529.
- Smart, J. S. (1968). 'Statistical properties of stream lengths', *Water Resources Research*, **4**, 1001–1014.
- Smart, J. S. (1969). 'Topological properties of channel networks', *Bulletin of the Geological Society of America*, **80**, 1757–1774.

- Smart, J. S. (1970). 'Use of topologic information in processing data for channel networks', *Water Resources Research*, **6**, 932-936.
- Smart, J. S. (1972). 'Channel networks', *Advances in Hydrosience*, **8**, V. T. Chow, editor. New York, Academic Press.
- Smart, J. S. (1973). 'The random model in fluvial geomorphology', *Proceedings of the Fourth Annual Geomorphology Symposium*, Marie Morisawa, editor, Binghamton, New York, p. 27-49.
- Werner, C. and Smart, J. S. (1973). 'Some new methods of topologic classification of channel networks', *Geographical Analysis*, **5**, 271-295.
- Werrity, A. (1972). 'The topology of stream networks', *Spatial Analysis in Geomorphology*, R. J. Chorley, editor, London, Methuen, 167-196.