EARTH SURFACE PROCESSES, VOL. 3, 129–170 (1978)

THE ANALYSIS OF DRAINAGE NETWORK COMPOSITION

J. S. SMART
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, U.S.A.

Received 15 December 1976
Revised 28 April 1977

SUMMARY

The Horton method of analyzing drainage network composition is reviewed, with the conclusion that it has not been very effective either in improving understanding or in developing useful methods of characterizing drainage basins. New methods which are based on the link rather than the Horton (or Strahler) stream are described. A number of detailed examples of the application of these new methods to the topologic and geometric properties of networks are provided. The results are compared with the predictions of the random model.

Data used in the analysis were obtained from 1:24,000 U.S.G.S. topographic maps of eastern Kentucky. Thirty drainage basins were selected and their channel networks were outlined first by the contour-crenulation (CC) method and then by another, more objective, method (SC) in which stream sources were identified by a quantitative slope criterion. The CC and SC samples comprise about 8,700 and 1,700 links, respectively.

The three most important results of the analysis are: (1) the channel networks are slightly but significantly more elongated than predicted by the random model, (2) there are fewer second magnitude links than predicted, and (3) the length distribution for interior links depends upon the kind of link (interior or exterior) joined downstream. These features are found in both CC and SC networks.

KEY WORDS Drainage networks Kentucky Link lengths

INTRODUCTION

The quantitative analysis of drainage network composition began with the well known paper by Horton (1945), and practically all of the early work in this field was based on the Horton or Strahler (1952) ordering procedures. Recent advances, however, beginning with Shreve (1966), have emphasized a viewpoint independent of ordering rules. New variables for characterizing network properties have been proposed, as well as new methods of data handling particularly suitable for arborescent networks. A theoretical model of drainage basin composition which provides a good approximation to natural networks has been developed. The model also, as suggested by Krumbein and Shreve (1970), can be used as a statistical standard for comparing and characterizing natural networks.

Although some or all of these new advances have now been employed in several papers, most notably in a comprehensive study by Jarvis (1975, 1976a, b, c), their intrinsic possibilities are far from being fully exploited. One minor but pleasant problem is that so much detailed information about network composition becomes available, that it is literally difficult to know where to start. The purpose of this paper is to provide a systematic set of examples of the power and efficiency of these new methods in extracting information about the planimetric features of channel networks from a collection of link-based data. In view of the exploratory nature of the investigation, some failures and dead ends will be reported along with the successes. A considerable amount of tabular data is presented for the benefit of those who may wish to check or expand the results presented here.

Before starting the discussion of the new methods, however, the results obtained with the Horton-Strahler approach are reviewed and assessed in the next section.

0360-1269/78/0203-0129 $01.00

THE HORTON ANALYSIS

In his famous paper on quantitative geomorphology of drainage basins, Horton (1945, p. 286) introduced the term *composition of the drainage net* with the statement.

> Composition implies the numbers and lengths of streams and tributaries of different sizes or orders, regardless of their pattern

The definition is thus specifically linked to Horton's ordering scheme. From the context of his paper, however, it appears that Horton's intention can be restated in more general and more modern terms by saying that drainage network (or basin) composition refers to the topologic and geometric properties of channel networks. It is in this sense that the term is used in this paper.

Horton's goal was to replace the qualitative descriptions of drainage basins (e.g., youthful, mature, well drained) with quantitative ones. To this end, he devised a method of classifying the segments of a channel network according to order and showed that the numbers and mean lengths of segments, or streams, of successive orders form approximate geometric progressions. Horton (1945, p. 295–300) then proposed to describe drainage basin composition by a set of physiographic factors derived with the aid of his stream ordering method. These procedures, now usually referred to as a 'Horton analysis', have been generally accepted and widely used in geomorphology and hydrology.

In a typical Horton analysis, the channel network is outlined on a map, the stream segments are ordered (now most commonly by Strahler's (1952, p. 1120) modification of Horton's method) and the stream numbers N_ω and mean stream lengths $\bar{L}_\omega$ are determined for the various orders ω. The geometric progression property is displayed by showing that plots of $\log N_\omega$ and $\log \bar{L}_\omega$ versus ω are approximately linear. Finally, the bifurcation ratio R_B and the stream length ratio R_L are derived as the negative and positive logarithms, respectively, of the best-fit straight lines to the plots. Other data which are sometimes collected include mean basin areas $\bar{A}_\omega$, mean slopes $\bar{S}_\omega$, and the corresponding ratios R_L and R_S. Detailed accounts of these procedures are given in many places in the literature (e.g., Smart (1972a, p. 307–10)). Figure 1 shows a Horton diagram constructed from data provided by Morisawa (1962).

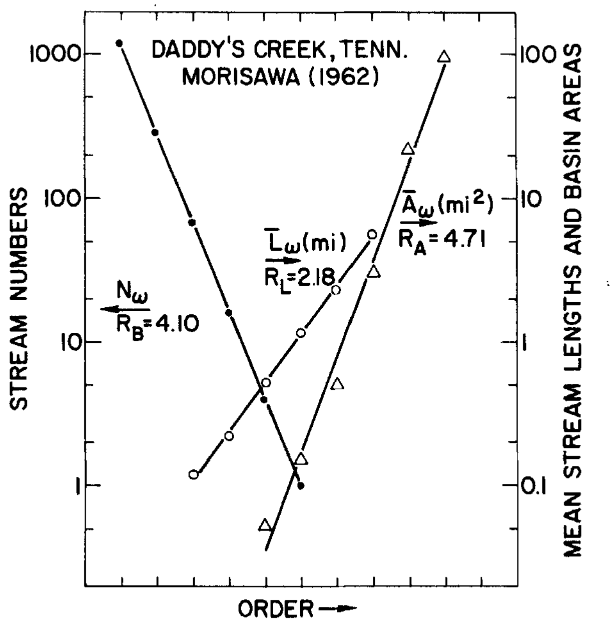


Figure 1. Horton diagrams for Daddy's Creek, Tenn. Data from Morisawa (1962) (reproduced by permission of Publications in Geomorphology, Inc.)

Horton (1945, p. 290) felt that his discoveries of the relations between stream numbers, mean stream lengths, and order were important because they represented 'laws that evolved from physical processes' and because of their practical applications. He was, however, rather vague about the nature of the practical applications and readers must infer his intentions from isolated claims and a few examples. The following paragraphs summarize my understanding of Horton's ideas about the application of the Horton analysis to drainage basin composition.

The first application suggested was the use of measured properties of a drainage basin to estimate unmeasured ones. Horton noted that if the five variables A(basin area), Ω(basin order), R_B, R_L, and $\bar{L}_1$ were known, many other properties of the stream system could be estimated. He also observed, however, that the labour required to obtain these five quantities was 'practically prohibitive', except for very small basins, and proposed to bypass much of this labour by also estimating these quantities from a relatively small number of measurements made on small-scale maps that did not show the low-order tributaries. Both steps in this procedure depend on the assumption that the geometric progressions are exact; thus the estimates are accurate only if all individual bifurcation ratios $N_\omega/N_{\omega-1}$ are close to R_B and all individual length ratios $\bar{L}_\omega/\bar{L}_{\omega-1}$, are close to R_L. Although Horton correctly recognized some of the limitations of his method, he does not seem to have been concerned about this particular point.

Horton (1945, p. 302) also believed that 'genetically similar' stream systems in similar geologic and climatic settings, should have nearly identical compositions. He did not develop this idea in any detail but the implication seems to be that a single set of physiographic parameters could characterize most, if not all, drainage basins in a given region and could distinguish them from drainage basins in another region with a different environment. Whatever Horton's exact ideas may have been, this suggestion is potentially the most important of his contributions. Unfortunately, the data which he presented in its behalf were scant and not very convincing, and there has been little if any later work which supported it. Early work by Strahler and his students led Strahler (1952, p. 1136; 1957, p. 914), to conclude that the bifurcation ratio was essentially independent of environment except in regions that have certain types of strong geological controls. Chorley (1957), in a morphometric study of three areas with widely different reliefs and drainage densities, found no significant differences in the bifurcation, length, or area ratios. Similar results were obtained by Woodruff (1964), Eyles (1968), and Shreve (1969, p. 413–4). In one exception to these generally negative results, Abrahams (1972) found a weak but significant (0·05 level) positive correlation of R_B with relief.

Horton's proposal to use information from small-scale maps to estimate the properties of the complete network has seen considerable use, e.g., Yang and Stall (1971). It was however, apparently not tested until Smart (1973) evaluated it with data on 38 fifth-order networks and found that it was both highly inaccurate and badly biased. The main contribution to the bias is a systematic deviation from the assumption that all individual bifurcation ratios are the same; instead, low-order ratios tend to be larger than high-order ones.

Thus the Horton analysis has proved unsuccessful in the two major applications that Horton envisaged for it. In geomorphic studies, there remains the rather modest possibility that the Horton parameters may be used simply to characterize and distinguish the individual network to which they apply. However, deficiencies in R_B and R_L as purely descriptive quantities were noticed by various early investigators. Schumm (1956, p. 603) and Maxwell (1960, p. 12) both observed that the plots of $\log N_\omega$ versus ω, from which R_B is derived, show small but systematic deviations from a straight line, with a tendency to be concave upward. In a comprehensive study of steam number data, Shreve (1966) confirmed the statistical significance of this effect, which is of course closely related to the trend of bifurcation ratio with order mentioned in the previous paragraph. Broscoe (1959, p. 5), Maxwell (1960, p. 23), and Bowden and Wallis (1964) all pointed out that the corresponding plots for stream lengths frequently show such large (but apparently unsystematic) deviations from a straight line that the representation of length data by a single average slope is of dubious value. Figure 2 shows a graphic example supplied by Smart (1973, p. 38).

More recently, Jarvis (1972) and Jarvis and Werritty (1975) have criticized the Horton analysis for

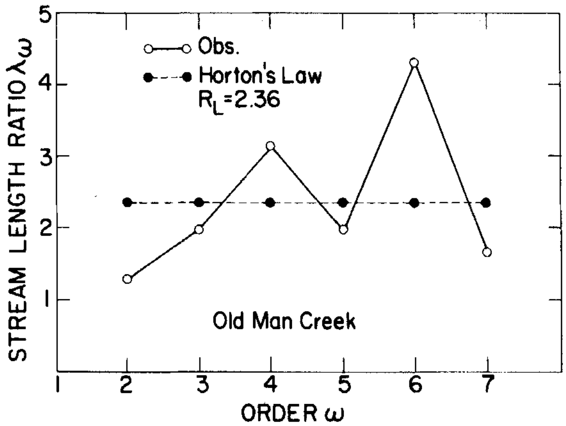


Figure 2. Observed stream length ratios $\lambda_\omega = \bar{L}_\omega/\bar{L}_{\omega-1}$ for Old Man Creek, Iowa. Data from Milling and Tuttle (1964); dotted line indicates Horton's law approximation (reproduced by permission of Publications in Geomorphology, Inc.)

the loss of topologic information inherent in the ordering procedure. In particular they noted that the possible topologically distinct channel networks for a given set of stream numbers usually encompass a wide range of values for basic topologic variables, such as diameter and mean exterior path length. For example, the set of networks with stream numbers 25, 6, 2, 1 have diameters ranging between 7 and 19, while the maximum possible range for all magnitude 25 channel networks is only 6–25.

In searching the literature it is difficult to find papers in which the descriptive function of the Horton analysis is employed in any substantive way. In such examples as do occur, the description is usually in terms of the mean link lengths and basin areas rather than stream numbers and ratios, e.g., Strahler (1958, p. 293–4). In many cases the results of the analysis are simply presented as a sort of assigned exercise, sometimes accompanied by a statement to the general effect that they demonstrate that drainage basins develop in a regular and orderly fashion. However reassuring this last claim may be, it is unfortunately open to challenge. Much of the regularity noted in the behaviour of the Horton parameters is a consequence of the rules of ordering and has little to do with the development of the drainage basin. For example, Shreve (1966, p. 30) showed that plots of N_ω versus ω must lie in a rather restricted region of the $N_\omega - \omega$ plane, and that no bifurcating network of any kind, whatever its origin, can depart indefinitely from Horton's geometric-series law.

Horton (1945, p. 286, 292) also felt that his methods would be useful in hydrology although his specific remarks are limited to the unelaborated comment that 'composition has a high degree of hydrologic significance', and to the suggestion that networks with a large value of R_L/R_B should have a large channel storage capacity. There have been several attempts by other workers to associate the dimensionless Horton parameters with the hydrologic properties of a watershed, but in the main they have been unsuccessful. The only exception that I have found occurs in two early and important papers by Anderson and Trobitz (1949) and Anderson (1949). In the first paper, the authors studied the relations between certain watershed variables and peak discharge and sediment deposition. Their analysis was based on observations on 40 drainage basins in the San Gabriel and San Bernardino Mountains of southern California during the major flood of 1938. They found that the physiographic variable $R_B R_S/R_L$ had a highly significant positive correlation with peak discharge. (This result was unfortunately improperly described by Werner and Smart (1973, p. 275).) In later work, Anderson (1949) extended the study to include storms of different sizes; he found that the effect of the physiographic variable was itself dependent on the storm intensity and antecedent precipitation. (This conclusion is actually reported (Anderson (1950)) in a discussion of the results of the original paper.) It should be noted, however, that because of the high correlation between R_B and R_L—unrecognized at that time—their effects on the physiographic variable probably tend to cancel, leaving it mainly dependent on the slope ratio R_S. This result is consistent with the known correlation of peak discharge with mainstream slope, e.g., Benson (1962).

Maxwell (1960) continued the study of the San Gabriel watersheds with a detailed geomorphic analysis. He found a weak but significant correlation between peak height and R_L but no significant correlation with either R_B or R_A. Similarly, Morisawa (1962) found that the bifurcation ratio had no significant influence on the peak discharge and mean annual runoff of drainage basins in the Appalachian Plateaus physiographic province. Thomas and Benson (1970) studied a large number of streamflow characteristics of watersheds tributary to the Missouri and Arkansas Rivers and found that R_B had no effect on any of them.

Leopold and Miller (1956) extended Horton's ideas by showing that the logarithms of certain hydrologic and hydraulic variables (such as discharge, channel width, sediment load) are approximate linear functions of basin order Ω. As noted by Leopold and Miller, the source of this behaviour lies in the fact that all of the quantities involved depend rather strongly on the size of the drainage basin. Moreover, the relation between a hydrologic variable X and area, which is the most common measure of basin size, can be expressed empirically as a power law of the general form

$$X \propto A^b$$

where b is some constant. This relation in turn implies that

$$\log X \propto \log A$$

and since Ω is also approximately proportional to $\log A$ (Horton (1945, p. 294)), the linear relation between $\log X$ and Ω follows immediately.

This application of stream ordering has been rather widely used for purposes of description and estimation. The principal benefit apparently is that it provides a simple and objective (though not entirely consistent) method of grouping drainage basins according to size. For much the same reasons, order has also been used to define populations of drainage basins for statistical geomorphic studies, e.g., Coates (1958). It is certainly not clear that order has any real advantage for such purposes over other size-related properties such as area and magnitude; an equally effective basis for classification and description could be obtained by dividing the logarithms of such quantities into appropriate ranges. Order is perhaps more easily measured than area but in practice area is such an important variable that it is usually measured for other reasons even when order is employed in part of the analysis. No doubt the fact that order is a simple integral measure works in its favour; attempts to create a similar kind of grouping by partitioning the logarithms of area or magnitude into somewhat arbitrary ranges may seem somehow more subjective and less satisfactory.

There is, however, one well known disadvantage in classifying stream networks by order. The wide range (theoretically infinite) of possible magnitudes for any order other than the first can introduce basins of anomalous size into an otherwise homogeneous group. For example, a third-order network of magnitude 47, say, almost certainly has an area and other size-related characteristics appropriate to fourth order, but because of an accident of topology it becomes an outlier in the third-order group.

Although it is not the purpose of this paper to encourage the continued use of ordering methods, I will suggest that this difficulty could be avoided by a simple change in the definition of order. Consider the sequence of magnitudes 1, 2, 3,... Let first-magnitude 'networks' be first order, as in the past. Assign the next four magnitudes (2–5) to second order, the next sixteen (6–21) to third order, and so on. The general rule is that order ω covers a range of $4^{\omega-1}$ magnitudes. Another way of specifying the sequences is that the maximum possible magnitude for order ω is $(4^{\omega}-1)/3$. This assignment of order, which depends only on network magnitude, clearly would produce more homogeneous groupings than the usual Strahler ordering, while at the same time retaining its general size characteristics. Also, the method does have some basis in theory; according to Shreve's (1966) random topology model, the range of magnitudes for each order is just that in which the corresponding Strahler order is the most probable one.

In conclusion, it seems fair to say that, after 30 years use, the benefits derived from the Horton analysis have been few and limited. In particular, it has provided neither improved understanding of drainage basin composition nor effective practical methods of describing and estimating basin characteristics. This failure is undoubtedly related to an observation made by several writers, e.g., Shreve (1966, p. 21), that the Horton parameters tend to be intimately related to each other but not to other watershed variables. In a sense, the ordering procedure erects an elaborate facade which hides the more fundamental properties of the network.

RECENT DEVELOPMENTS

The critics of the Horton approach have naturally suggested alternative procedures which they feel will afford improved methods of analysis of drainage network composition. These suggestions can be divided into three categories: (1) new geomorphic variables to replace those derived from the Horton analysis, (2) new methods of data treatment, and (3) a new model of channel network composition to supplant Horton's laws. Although the histories of these three developments are closely interrelated, it should be noted that they are logically independent of each other. That is, it is possible to make use of the new variables without adopting either the new methods of data treatment or the random model, and so on. In this paper, however, we will, as stated in the introduction, employ all of them as needed to clarify and explain the composition of natural networks.

The topologic terms used in this paper are taken mainly from Shreve (1966, p. 20, 27; 1967, p. 178–9) with a few additional suggestions from Werner and Smart (1973). A reasonably complete set of definitions is given by Smart and Werner (1976).

Link-associated variables

In a Horton analysis, the variables of interest are all related to properties of Strahler (or Horton) streams. Many recent workers, however, have followed Shreve's (1966) suggestion that the link rather than the Strahler stream segment be regarded as the basic unit of channel network composition. As illustrated in Figure 3, the choice of a link in a network also defines a path and a subnetwork. Thus, three different kinds of geomorphic variables may be associated with links. We shall refer to them as link, path, and network variables, according to whether they are identified with one and only one link, with the path between the link and the outlet, or with the subnetwork defined by the link. Some examples are given in Table I. Path variables are in general functions of link variables, and network variables may be functions of either path or link variables. For example, the mainstream length of a network is the largest geometric path length, which is in turn the sum of the individual link lengths;

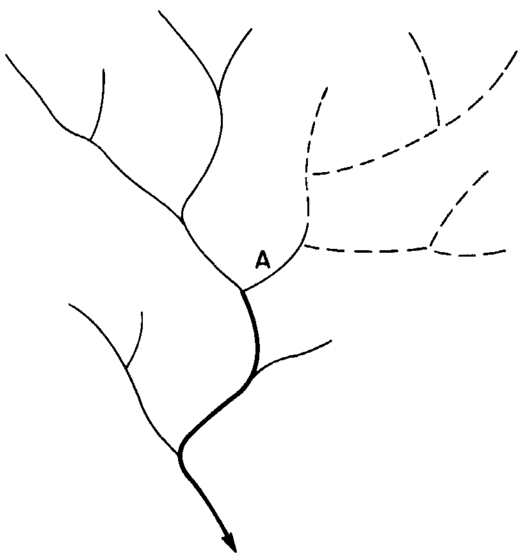


Figure 3. Channel network of magnitude 12. Topologic vector is 00110000110110010110111. Link A defines a path consisting of itself and a the heavy-line segment and a subnetwork consisting of itself and the broken-line segment

Table I. Link-associated variables

Type of variable	Examples
Link	Length; associated drainage area; orientation
Path	Topologic path length; flowpath length; sinuosity
Network	Magnitude; basin area; total channel length; drainage density; link frequency.

again, the basin area is the sum of the individual link drainage areas. The use of path variables (and the network variables derived from them) in drainage network analysis was suggested by Jarvis (1972) and a theoretical background was supplied by Werner and Smart (1973).

The principal advantage claimed for these new variables is that they are more closely related to important geomorphic and hydrologic properties than are the Horton variables (Werner and Smart (1973, p. 284–5); Jarvis and Werritty (1975)). As an example, the mean time of travel is a useful concept in hydrologic analysis. It is immediately clear that the mean flowpath length would provide a good approximate measure of the property for most networks. None of the Horton parameters, on the other hand, have any obvious relation to mean travel time.

Link data vectors

The labour of measurement and computation involved in handling this potentially large number of new variables can become prohibitive unless corresponding new procedures are devised for carrying it out. Smart (1970) proposed a method for determining a large number of link-associated variables from a relatively small number of link measurements. Computer processing is required, however, to make efficient use of it. Smart's paper dealt specifically with information retrieval, and the general ideas are restated here in a form more applicable to drainage network analysis.

At the very least, two kinds of link variables, classification (exterior or interior) and length, must be measured for a meaningful study of drainage basin composition. Measurements of other link properties, such as associated drainage area and slope, greatly increase the information that can be obtained. The data on each variable is recorded in a string, or vector, with the appropriate value for a given link occupying the same position in each vector. The order in which the links occur in the vectors is determined by a rule proposed by Shreve (1967): starting at the outlet, traverse the network by making a left turn at each junction and reversing direction at each source; the datum for each link is recorded the first time it is traversed. The data vectors for a network of magnitude n will each have $2n - 1$ elements, n for exterior links and $n - 1$ for interior links.

The convention used by Smart (1970) for link classification was to identify exterior links by ones and interior links by zeros. (See Figure 3 for an example.) The resulting vector was called 'the binary string', but topologic vector might be a better term, since it contains complete topologic information about the network. It should be noted that not all sequences of n ones and $n - 1$ zeros specify a network; the necessary and sufficient condition is that the number of ones cannot exceed the number of zeros until the last element is reached (Shreve (1967, p. 183)).

Once the basic data vectors have been constructed, other link, path and network variables (including the Horton parameters) can be determined with the aid of various simple algorithms. To compare the measurement labour involved with that for a Horton analysis, note that these procedures require one measurement of each kind (length, area, etc.) for each link in the network while the Horton analysis requires one measurement of each kind for each Strahler stream. Since a network of magnitude n contains $2n - 1$ links and approximately $4n/3$ Strahler streams (Shreve (1967, p. 185)), the number of measurements required for the link data method is about 50 per cent more than for a Horton analysis. The justification for this extra labour is of course that it greatly increases the amount of information that can be obtained about drainage basin composition. In addition to the convenient computation of many variables, use of data vectors also permits sorting the subnetworks according to various properties such as magnitude, area, diameter, mainstream length, etc. Examples of this kind of data treatment are reported by Jarvis (1975) and Werner and Smart (1976).

Smart (1970) has given general algorithms for obtaining path and network variables from the link data vectors. The computations can frequently be made more efficient, however, by using specific algorithms for specific purposes. For example, Jarvis (1975) has described a convenient method of determining the magnitude vector by a repeated set of operations on the topologic vector. As the ones which identify exterior links in the topologic vectors are also the magnitudes of such links, the function of the algorithm is simply to assign magnitudes to the interior links. The required steps are as follows.

1. Locate the position, k, of the last zero from the left end of the topologic vector.
2. Replace the zero by the sum of the $(k + 1)$th and $(k + 2\mu)$th elements, where μ is the $(k + 1)$th element
3. Repeat steps 1 and 2 until all zeros have been replaced. The result is the magnitude vector.

The $(k + 1)$th link is of course the left fork of the kth link; if it has magnitude μ, a subnetwork of $2\mu - 1$ links must be traversed before the right fork is reached at the $(k + 2\mu)$th link. Taking the rightmost zero each time insures that the subnetwork magnitudes are evaluated 'from the top down' with no gaps in assignment.

The magnitude vector is often used in algorithms for determining other variables. An example is the algorithm for the topologic path length, which proceeds by repeated operations on a vector P whose initial value consists of a one followed by $2n$ zeros.

1. Locate the position, j, of the first zero from the left end of P.
2. Assign the value $1 + p$ to the jth and $(j + 2\mu - 1)$th elements, where p is the $(j - 1)$th element of P and μ is the jth element of the magnitude vector.
3. Repeat steps 1 and 2 until all zeros have been replaced. The result is the path length vector.

In this algorithm, the outlet vector is assigned a path length of one, and all other values are then determined by the rule that both forks of a link with path length p have path length $1 + p$. The magnitude vector is used to locate the right fork. Algorithms for computing geometric path and network variables from the link length and other geometric data vectors also employ the magnitude vector in a straightforward extension of the method described above.

Although these general procedures could be used in almost any conceivable type of drainage network analysis, they are not necessarily the most convenient for all applications. For example, in the field of hydrograph analysis, the methods devised by Surkan (1969, 1974), Surkan and Kelton (1974), and Kirkby (1976) specifically for that purpose appear to be more convenient and efficient than the link data method.

The random model

The earliest suggestion of the importance of random processes in the development of drainage basins came from Leopold and Langbein (1962). Arguments emphasizing the necessity of a probabilistic approach to geomorphic phenomena in general and to channel network composition in particular have been put forth by Scheidegger and Langbein (1966), Krumbein and Shreve (1970, p. 40), and Shreve (1975). The random model of drainage basin composition is a specific quantitative formulation of these general ideas. The model had its source in a paper by Shreve (1966) and was developed by Shreve, Smart, Werner and others. A concise account of the status of the model and references to most of the preceding papers are given by Shreve (1975).

The content of the model can best be expressed in two basic postulates (Smart (1973, p. 31–2)).

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

The postulates are generally supported by observations on natural mature drainage basins although deviations have been noted for certain specific areas or in certain specific details. The principal success of the model lies in the fact that many features of channel network composition, including Horton's laws, can be explained quantitatively and without adjustable parameters. Examples are listed by Smart (1973) and Shreve (1975).

These two assumptions are somewhat reminiscent of Horton's first two laws, since the first postulate deals with topologic properties and the second postulate with channel length properties. Although their content is much greater than that of Horton's laws, it may be noted that they do not afford a complete account of drainage basin composition. Just as a law of basin areas (Schumm (1956)) was added to Horton's original pair, a third postulate concerning areal properties is needed to extend the random model. The most obvious possibility is a statement about exterior and interior link drainage areas analogous to the postulate about exterior and interior link lengths. Such an assumption has been employed in various ways by Shreve (1969, 1974), Werner and Smart (1973), and Smart and Werner (1976). One new feature which appears here is the rather high correlation between the link lengths and the corresponding drainage areas. This point was not important in the Werner-Smart applications and was handled empirically by Shreve (1974) in his Monte Carlo studies, but it clearly must be taken into account in any attempt to extend the model. Although the relation between link lengths and areas has been the subject of a limited amount of theoretical analysis (Shreve (1974); Smart (1976)), it seems advisable to postpone the construction of the third postulate until our understanding has improved considerably. Finally, the model could be completed and its value greatly enhanced by the addition of a fourth postulate concerning channel slopes and their relation to the planimetric properties of basins; see, for example, a recent contribution by Flint (1976).

Smart and Werner (1976) demonstrated the effectiveness of the random model in estimating the values of geomorphic variables and in explaining and predicting geomorphic relationships. The use of link-associated variables and link data vectors was mentioned but not emphasized. This paper, which is to some extent a complement of the previous one, illustrates the application of these two new developments to the analysis of drainage basin composition and employs the random model as a statistical standard against which natural drainage basins are compared. Also, the considerable body of network data collected by Smart and Werner is used as a basis for this study. The collection and handling of this data was briefly described in that paper but a more detailed account is given in the next section.

COLLECTION AND HANDLING OF DATA

The data were obtained by map studies of drainage networks in an area of eastern Kentucky that includes the watersheds of Tug Fork and Levisa Fork of the Big Sandy River and North Fork and Middle Fork of the Kentucky River. This area, which is called by Price *et al.* (1962) the Eastern Coal Fields Region, lies in the Kanawha section of the Appalachian Plateaus province. The Tug Fork drainage basin 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 broad-crested 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 geological controls.

Comparison of maps and consultation of the literature on the geology and physiography of Kentucky, especially Price *et al.* (1962) and McFarlan (1943), indicate this description should also apply to the somewhat more extensive region covered in this study. The temperature and precipitation records do not suggest any significant differences in climate throughout the area. The outlet altitudes of the networks chosen for study lie in the range 182–244 m (600–800 ft). All of the U.S.G.S. topographic maps of the region were prepared by photogrammetric methods from aerial photographs.

For both this and the Smart and Werner investigations it was desirable that the sample of networks include a considerable range of sizes. Target magnitudes for the sample were obtained by randomly drawing 24 numbers from a uniform distribution of integers between 40 and 270. Then a set of 24 networks with magnitudes matching the target magnitudes as closely as possible were identified on the 1:24,000 U.S.G.S. topographic maps of the Tug Fork and Levisa Fork areas (Offut, Inez, Kermit, Lancer, Thomas, Varney and Williamson quadrangles). The location of outlet links was thus not a purely random selection but is believed to be unbiased. Six networks from the North and Middle Fork basins (Canoe, Haddix, and Noble quadrangles) that were originally selected for another purpose were later added to the sample upon finding that the Mann-Whitney U test (Siegel (1956, p. 116)) for drainage density and mean link lengths did not reject the hypothesis that the two groups came from the same population. The two localities are separated by about 80 km. No studies were made in the intervening region, which includes the headwaters of the Licking River, because those maps have a different contour interval—20 ft, as opposed to 40 ft in the ones that were used. Measurements on basins that extended across the boundary of two adjacent quadrangles with different contour intervals showed small but significant differences in drainage density and link frequency for the two parts.

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 segments. In addition to the link data, the total drainage area, total exterior link drainage area, and total interior link drainage area were measured for each basin with a polar planimeter.

Accuracy and reproducibility of map data in studies such as this are of paramount importance. Ideally, it is desirable that each source located on the map correspond to a natural source and that no 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 information can still be obtained if the sources and channels are identified by reasonably objective map criteria and can be reproduced with reasonable accuracy by different operators. Krumbein and Shreve (1970, p. 1938) have made a detailed investigation of operator variation in map measurement for this same region. Smart and Werner (1976) 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. They found that the operator variation in determining the number of exterior links was almost always less than 5 per cent. This point 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. They 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 links. The differences between two observers for mean link length of samples of size 100 were typically about 10 per cent for exterior links and 2 or 3 per cent for interior links. All data used in this paper were collected by one person, Thomas McElroy, a student at the Department of Geological Sciences, SUNY, Binghamton, New York.

The input data for the analysis consisted of the 30 topologic vectors, the 30 link length vectors, and the basin areas. The area data was entered as three vectors, (A, A_e, and A_i), each of length 30, with the individual networks appearing in the same sequence in all three cases. Data vectors of this type, which record properties of the complete networks, will be called watershed vectors, the name being specifically chosen to avoid excessive and confusing use of the terms network and subnetwork. The 60 link data vectors and three watershed data vectors were stored in the APLSV system of an IBM 360/91 computer. Table IIb lists 30 variables that were computed with algorithms of the type described in the previous section. The CPU time required was less than one minute. This sample of output data (90 link vectors and 27 watershed vectors) by no means exhausts all of the possibilities but it is sufficient to provide a sound basis for a study of drainage basin composition.

The 30 networks contain a total of 4,377 exterior links and 4,347 interior links. Network magnitudes range from 41 to 267 and basin areas from 1·84 to 13·68 km^2 (0·71–5.29 m^2). Some of the more important

Table IIa. Input data

Vector type	Variable	Symbol
L	Link classification (topologic vector)	
	Link length	l
W	Basin area	A
	Exterior basin area	A_e
	Interior basin area	A_i

Table IIb. Output data

L	Magnitude	μ
	Topologic path length	j
	Geometric path length	J
W	Magnitude	n
	Total channel length	L
	Drainage density	D
	Exterior drainage density	D_e
	Interior drainage density	D_i
	Link frequency	F
	Melton parameter (D^2/F)	K
	Exterior K	K_e
	Interior K	K_i
	Exterior path length properties	
	Maximum (diameter)	d
	Total	p_e
	Mean	$\bar{j}_e$
	Standard deviation	s_e
	Coefficient of variation	C_e
W	Flowpath length properties	
	Maximum (mainstream length)	L'
	Total	P_E
	Mean	$\bar{J}_E$
	Standard deviation	S_E
	Coefficient of variation	C_E
W	Mean exterior link length	$\bar{l}_e$
	Mean interior link length	$\bar{l}_i$
	Link length ratio	λ
	Mean exterior link drainage area	$\bar{a}_e$
	Mean interior link drainage area	$\bar{a}_i$
	Link drainage area ratio	α
W	Topologic shape factor ($d^2/2n - 1$)	β_T
	Geometric shape factor $(L')^2/A$	β_G

* L = link data vectors.
W = watershed data vectors.

properties of the individual basins are listed in Table III. Other information, including identification and location of the networks, can be obtained upon request. Table IV gives descriptive statistics on some variables that are expected to be characteristic of the local environment and relatively independent of basin size.

Table III. Properties of individual drainage basins

Network No.	n	d	p_e	A (km^2)	L (km)	D (km^{-1})	L' (km)	l_e (km)	l_i (km)
1	267	59	8,585	12·29	69·3	5·64	6·52	0·145	0·115
2	200	54	5,825	9·27	53·1	5·73	7·06	0·152	0·114
3	241	37	5,615	10·60	59·7	5·64	4·30	0·144	0·104
4	220	53	6,347	12·62	60·5	4·80	5·97	0·154	0·122
5	208	43	5,335	9·28	54·6	5·88	4·70	0·162	0·101
6	199	50	6,083	9·92	55·0	5·55	6·24	0·159	0·118
7	177	55	5,423	10·29	57·8	5·61	7·35	0·206	0·121
8	175	50	4,968	8·02	43·9	5·48	5·20	0·144	0·108
9	156	45	3,783	6·47	42·4	6·56	4·22	0·167	0·105
10	150	37	3,238	5·49	35·9	6·54	3·88	0·142	0·098
11	141	44	3,486	6·08	34·6	5·69	5·20	0·148	0·098
12	134	38	2,942	6·23	35·8	5·75	3·91	0·157	0·111
13	111	37	2,422	7·42	37·7	5·07	5·07	0·207	0·133
14	86	24	1,167	5·93	28·9	4·88	3·48	0·184	0·155
15	169	35	3,297	5·64	42·1	7·46	3·15	0·157	0·093
16	72	31	1,279	3·56	19·9	5·59	3·56	0·184	0·093
17	121	39	2,838	3·47	27·8	7·99	3·03	0·150	0·082
18	53	19	612	1·83	11·8	6·40	1·72	0·132	0·092
19	79	36	1,760	2·23	15·9	7·09	2·93	0·123	0·078
20	60	18	694	2·02	13·6	6·72	2·45	0·123	0·105
21	111	27	2,036	5·77	28·0	4·85	3·14	0·142	0·111
22	41	18	490	1·92	9·7	5·04	2·14	0·127	0·113
23	79	24	1,218	4·31	24·0	5·57	3·57	0·163	0·143
24	104	41	2,724	4·88	27·5	5·64	4·14	0·166	0·100
25	147	48	4,218	7·47	40·7	5·45	5·78	0·159	0·119
26	211	43	5,075	13·68	65·2	4·76	5·84	0·173	0·136
27	139	35	3,002	7·34	37·1	5·05	3·94	0·141	0·127
28	65	28	1,092	3·71	20·4	5·51	3·88	0·178	0·139
29	224	62	7,451	9·68	59·6	6·16	6·73	0·162	0·104
30	237	62	8,137	10·17	66·5	6·54	5·78	0·175	0·106

Table IV. Descriptive statistics of some geomorphic properties

	D (km^{-1})	F (km^{-2})	K	λ	α
Mean	5·82	44·3	0·776	1·43	1·34
Standard deviation	0·78	10·3	0·089	0·21	0·28
Coef. of variation	0·134	0·232	0·115	0·148	0·208
Maximum	7·99	70·3	0·933	1·97	1·98
Minimum	4·76	28·8	0·603	1·11	0·80

TOPOLOGIC ANALYSIS

One consequence of the random model is that the planimetric properties of stream networks are strongly influenced by the underlying topologic structure (Shreve (1975)). This prediction is supported by the results of various investigations, particularly that of Werner and Smart (1976). The first step in our analysis of channel network composition, then, is an attempt to gain some insight into the topologic properties of the networks.

Diameter distributions

Next to magnitude, the most important topologic property of a channel network is the diameter. For networks of magnitude n, the possible values of the diameter are the integers between $2 + [\log_2(n-1)]$ and n, inclusive. (Here $[x]$ means the largest integer $\leqslant x$). Figures 4a and 4b give examples of channel networks having the maximum and minimum diameters, respectively, for $n = 8$. The diameter is clearly a measure of the longitudinal extent of the network (Jarvis (1972)) just as its geometric analogue, the mainstream length, is a similar measure of the drainage basin. Another topologic variable with somewhat the same properties is the mean exterior path length $\bar{j}_e = p_e/n$. The networks of Figures 4a and 4b also have extreme values of $\bar{j}_e$, but it should be noted that for $n \geqslant 9$ not all networks of minimum diameter also have minimum mean exterior path length. Jarvis and Werritty (1975) and Jarvis (1975) have designed networks as elongated or compact according to whether the values of d and $\bar{j}_e$ were greater or less than the expected values from the random model. This dual usage is not completely consistent, however, since it is rather easy to find examples of pairs of networks of the same magnitude with one having the larger diameter and the other the larger mean exterior path length; one such example is given in Figures 4c and 4d. In this paper, the diameter will be used as a measure of elongation.

The diameter distributions for subnetworks of a given magnitude were determined and compared with the predictions of the random model. The observed distributions were obtained by using the magnitude and path length algorithms previously described to locate subnetworks of the appropriate magnitude and finding their maximum exterior path lengths. A number of methods are available for computing the random model distributions for diameters; two methods are described by Werner and Smart (1973, p. 279 and p. 283) and one by Shreve (1974, p. 1169–70). Results were obtained for magnitudes in the range 4–20, and the observed and expected frequencies for magnitudes 4–12, along with the results of χ^2 goodness-of-fit tests, are shown in Table V. As the sample sizes become rather small at the largest magnitudes, it was necessary to combine some of the diameter categories in carrying out the χ^2 test. In order to make the test as sensitive as possible for extreme values, I have, following the recommendations of Cochran (1954), allowed expected values as low as 1, as long as not more than 20 per cent of the categories have expected values less than 5.

The hypothesis that the diameter distributions are those of the random model can be accepted at the 0·05 level in 7 of the 9 cases but is rejected for magnitudes 6 and 7. The results shown in Table

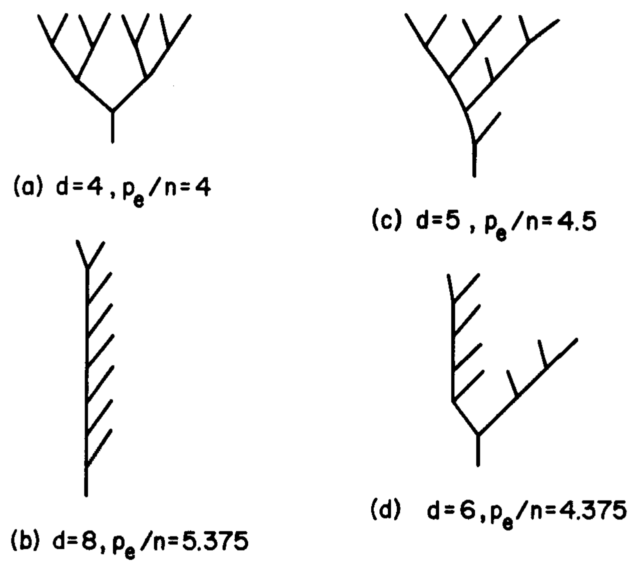


Figure 4. Channel networks of magnitude 8; (a) smallest d and p_e/n, (b) largest d and p_e/n, (c) and (d) example in which the relative sizes of d and p_e/n are interchanged

Table V. Theoretical and observed diameter distributions

<table>
<tr><th></th><th>Obs.</th><th>Exp.</th><th>Obs.</th><th>Exp.</th><th>Obs.</th><th>Exp.</th></tr>
<tr><td>d μ</td><td colspan="2">4</td><td colspan="2">5</td><td colspan="2">6</td></tr>
<tr><td>3</td><td>52</td><td>63·4</td><td></td><td></td><td></td><td></td></tr>
<tr><td>4</td><td>265</td><td>253·6</td><td>90</td><td>93·4</td><td>12</td><td>25·9</td></tr>
<tr><td>5</td><td></td><td></td><td>128</td><td>124·6</td><td>92</td><td>86·2</td></tr>
<tr><td>6</td><td></td><td></td><td></td><td></td><td>77</td><td>68·9</td></tr>
<tr><td></td><td>317[a]</td><td>2·56[b]</td><td>218</td><td>0·22</td><td>181</td><td>8·76</td></tr>
<tr><td></td><td></td><td>3·84[c]</td><td></td><td>3·84</td><td></td><td>5·99</td></tr>
<tr><td>d μ</td><td colspan="2">7</td><td colspan="2">8</td><td colspan="2">9</td></tr>
<tr><td>4</td><td>4</td><td>4·7</td><td rowspan="2">18</td><td rowspan="2">21·2</td><td></td><td></td></tr>
<tr><td>5</td><td>32</td><td>47·0</td><td>3</td><td>6·6</td></tr>
<tr><td>6</td><td>68</td><td>65·8</td><td>42</td><td>46·8</td><td>24</td><td>26·6</td></tr>
<tr><td>7</td><td>51</td><td>37·6</td><td>46</td><td>44·3</td><td>32</td><td>33·9</td></tr>
<tr><td>8</td><td></td><td></td><td>26</td><td>19·7</td><td>28</td><td>24·9</td></tr>
<tr><td>9</td><td></td><td></td><td></td><td></td><td>14</td><td>9·0</td></tr>
<tr><td></td><td>155</td><td>9·75</td><td>132</td><td>3·06</td><td>101</td><td>5·46</td></tr>
<tr><td></td><td></td><td>7·81</td><td></td><td>7·81</td><td></td><td>9·49</td></tr>
<tr><td>d μ</td><td colspan="2">10</td><td colspan="2">11</td><td colspan="2">12</td></tr>
<tr><td>5
6</td><td>7</td><td>15·8</td><td>2</td><td>8·9</td><td>1</td><td>3·8</td></tr>
<tr><td>7</td><td>23</td><td>23·7</td><td>14</td><td>19·3</td><td>6</td><td>12·0</td></tr>
<tr><td>8</td><td>27</td><td>22·6</td><td>28</td><td>22·6</td><td>18</td><td>16·9</td></tr>
<tr><td>9</td><td>18</td><td>13·7</td><td>21</td><td>17·7</td><td>16</td><td>15·8</td></tr>
<tr><td>10</td><td>5</td><td>4·2</td><td>13</td><td>9·2</td><td>14</td><td>10·6</td></tr>
<tr><td>11</td><td></td><td></td><td>2</td><td>2·4</td><td rowspan="2">10</td><td rowspan="2">5·9</td></tr>
<tr><td>12</td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>80</td><td>7·24</td><td>80</td><td>10·42</td><td>65</td><td>8·99</td></tr>
<tr><td></td><td></td><td>9·49</td><td></td><td>11·07</td><td></td><td>11·07</td></tr>
</table>

a. Sample size.
b. Chi-square statistic, X^2.
c. Critical value for 0·05 level of significance.

V are of course not independent; approximately half of the networks of magnitude μ combine with exterior links to form networks of magnitude $\mu + 1$. Thus it is perhaps not surprising that the two rejections occur in sequence. Similarly, for the data not shown in the table, the random model hypothesis is rejected for magnitudes 13 and 14.

These results might be considered as fair support for the random model. However, in such sets of χ^2 tests it is important to look for systematic patterns in the deviations between expected and observed values. One such pattern is noted almost immediately. For all magnitudes, the smallest diameters are less abundant than predicted and the largest diameters are more abundant than predicted. One minor exception occurs at magnitude 11, where the observed and expected frequencies for diameter 11 are 2 and 2·4, respectively. This pattern also holds for magnitudes 13–20. Thus the Kentucky networks of magnitudes 4–20 tend to be somewhat more elongated than those drawn from a topologically random population.

This conclusion is consistent with observations by Krumbein and Shreve (1970) on networks in the same area. For 153 magnitude 5 networks, they found 51 with diameter 4 and 102 with diameter 5, compared with the expected values of 65·6 and 87·4, respectively. They also noted that a sample of 90 third-order networks did not appear to be drawn from a topologically random population because they had too many first-order side tributaries or, in the terminology of this paper, were too elongated.

The question then arises as to whether this effect persists for larger networks. In order to test this possibility, the diameter distributions for given magnitude were computed up to $\mu = 200$, at which point the computing time required by the APLSV system began to be excessive. For large magnitudes, the sample sizes are too small for χ^2 tests for each magnitude and some other method must be used. The best procedure seems to be one devised by Shreve (1966) for combining small samples to see whether they could have been drawn from a family of distributions with the same form but different parameters. In this application of Shreve's test, the diameter distribution functions for each magnitude were divided into quartiles and each diameter observation was assigned to the appropriate quartile. If the probability range for a given diameter bracketed a quartile boundary, the observation was divided proportionally between the two quartiles. Under the random model hypothesis, the expected frequencies for the four classes are equal.

A sample for the test was selected from the 30 networks by choosing all networks with magnitudes in the range 100–200, and for networks with $n > 120$, choosing the largest subnetwork with magnitude less than or equal to 120. This process provided a sample of 22 networks (or subnetworks) with a magnitude range of 81–120 and a mean of 108. A similar procedure designed for larger networks produced a sample of 18 with a magnitude range of 134–200 and mean of 167. These groups are identified as I and II, respectively, in Table VI, where Q_1, Q_2, Q_3, and Q_4 are the quartile values listed in order of increasing diameter. Although the combined sample sizes are still rather small, it appears that these larger networks are also elongated by random model standards. The random model hypothesis is not rejected at the 0·05 level for Group I but the most noticeable deviation is a deficiency in the smallest diameter category. Group II displays the complete pattern of too few small diameters and too many large ones.

Returning to the results shown in Table V, we note that the distributions for magnitudes 4 and 5 agree well with the random model, while for $\mu = 6$, as previously mentioned, the fit is quite poor. This observation poses a minor mystery since most sixth-magnitude networks contain either a fourth- or fifth-magnitude subnetwork or both. It is clear that the main source of deviation for $\mu = 6$ is the lack of diameter 4 networks, since the combination of 12 observed and 25·9 expected contributes 7·46 to the chi-square statistic $X^2 = 8.76$. An obvious speculation is that the discrepancy is also related to the distribution of magnitude 3 networks, and we shall explore this possibility in some detail.

Some networks with $\mu = 6$ and $d = 4$ contain third-magnitude subnetworks and some do not. Neither the path length nor the path number classifications (Werner and Smart (1973, Table 1)) will resolve this dichotomy and it is necessary to go to the ambilateral classification (Smart (1968, p. 1760)). The results are shown in Table VII. For networks of magnitude less than 9, the ambilateral classes can be specified by giving the link magnitudes and this method has been used to identify them in Table VII. A network of magnitude n always has n links of magnitude 1, one link of magnitude n, and at least one link of magnitude 2. Consequently, the values μ of the $n - 3$ remaining interior link magnitudes are sufficient for identification. Figure 5 shows the ambilateral classes associated with the μ values.

The diameter 4 class divides into the 233 and 224 ambilateral classes and we see that it is indeed the former, in which two third-magnitude subnetworks combine, that produces the major part of the disagreement. Only five 233 subnetworks are observed, although 17·2 are expected and this deviation contributes 8·69 to the total χ^2 statistic of 11·67. The relative abundances of those classes which contain a single third-magnitude subnetwork (234, 235, and 345) are not very different from the random model predictions.

Table VI. Quartile analysis

Variable	Exp.	Q_1	Q_2	Q_3	Q_4	X^2
Diameter (I)	5·5	1·76	6·95	7·73	5·57	3·83
Diameter (II)	4·5	0·37	5·33	3·41	8·89	8·49*
$m\,(\mu = 3)$	7·5	15·19	5·54	4·23	5·13	10·39*
$m\,(\mu = 4)$	7·5	9·08	12·09	6·03	2·80	6·37

* Rejected at 0·05 level.

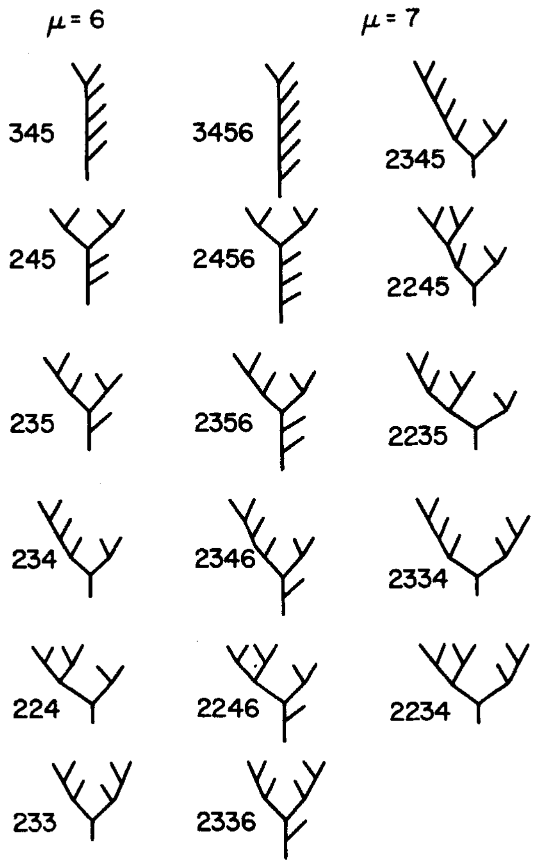


Figure 5. Ambilateral classes for magnitudes 6 and 7. Numbers indicate interior link magnitudes excluding 1, 2, and μ. Networks in middle column are formed by adding an exterior link at outlet of networks in left-hand column

Table VII also gives the ambilateral class distributions for $\mu = 7$ and is constructed so that each class for $\mu = 6$ is on the same line as its topologic descendant for $\mu = 7$. That is, the seventh-magnitude network is obtained by adding an exterior link at the outlet of the sixth-magnitude network. The shortage of 3–3 pairs is still noticeable but its effect is diluted by the addition of 5 more classes.

Table VII. Distribution of ambilateral classes

	$\mu = 6$				$\mu = 7$		
μ_i	d	Obs.	Exp.	μ_i	d	Obs.	Exp.
345	6	77	69·0	3456	7	51	37·6
245	5	16	17·2	2456	6	10	9·4
235	5	34	34·5	2356	6	16	18·8
234	5	42	34·5	2346	6	24	18·8
224	4	7	8·6	2246	5	5	4·7
233	4	5	17·2	2336	5	3	9·4
				2345	6	18	18·8
				2245	5	5	4·7
				2235	5	5	9·4
				2334	5	14	18·8
				2234	4	4	4·7
		181[a]	11·67[b]			155	14·50

a. Sample size.
b. Chi-square statistic X^2.

Another feature of the magnitude 7 distribution which contributes about as much to the χ^2 statistic is the overabundance of second-order, or 3456, networks (also noticeable in Table V). Now all class 3456 networks must come from class 345. If we ignore the small difference between number of exterior links and number of interior links in the total sample, then the expected number of class 3456 networks generated from the 77 class 345 networks is 38·5, compared with the observed value of 51. Thus, the excess of class 3456 is not due to any great overabundance of class 345, but rather that for some unknown reason, about 2/3 of class 345 combines with exterior links and only 1/3 with interior links.

Magnitude distributions

This point suggests another line of investigation. Let N_μ be the number of subnetworks (or links) of magnitude μ in our sample. Then

$$N_\mu = E_\mu + I_\mu, \quad \mu = 1, 2, 3, \ldots, 267. \tag{1}$$

where E_μ and I_μ are the numbers of subnetworks that join exterior and interior links, respectively, at their downstream end. Also

$$N_2 = E_1/2 \text{ and } N_3 = E_2 \tag{2}$$

since second-magnitude links can be formed only by combination of two exterior links and third-magnitude links only by combination of a second-magnitude link and an exterior link. There are other relations, both equalities and inequalities, which the N_μ, E_μ, and I_μ must satisfy in general, but their specific values for a given set of networks are largely determined by the topologic structure of the networks. Theoretical statistical distributions can of course be computed using the random model.

An algorithm to determine E_μ was devised and executed for $\mu \leqslant 12$, and some results concerning N_μ and E_μ are shown in Table VIII. Two distinct but related variables, N_μ, the absolute abundance of subnetworks of magnitudes μ, and E_μ/N_μ, the fraction of such subnetworks which combine with exterior links, are compared with the predictions of the random model for this specific sample.

A good approximation to the expected values of N_μ given by the random model can be obtained by computing the expected number of links of magnitude μ in a sample of 8,724 links of an infinite topologically random channel network. (Thus the approximation consists of ignoring the fact that the actual sample is a combination of data from 30 distinct networks.) Shreve (1967, p. 181) gives the probability that a link drawn at random will have magnitude μ.

$$v(\mu) = 2^{-(2\mu-1)} N(\mu) \tag{3}$$

Table VIII. N_μ and E_μ

μ	N_μ Obs.	N_μ Exp.	N_μ(obs.)/N_μ(exp.)	Simulation Results[a]	E_μ Obs.	E_μ/N_μ Obs.	z[b]
1	4,377				2,086	0·477	−3·10*
2	1,043	1,094·3	0·95	1	469	0·450	−3·25*
3	469	547·1	0·86	0	265	0·565	2·82*
4	317	342·0	0·93	1	160	0·505	0·11
5	218	239·4	0·91	3	127	0·583	2·37*
6	181	179·5	1·01	46	109	0·602	2·68*
7	155	141·1	1·10	3	91	0·587	2·09*
8	132	114·6	1·15	1	72	0·545	0·96
9	101	95·5	1·06	15	52	0·515	0·20
10	80	81·2	0·99	48	43	0·538	0·56
11	80	70·1	1·14	3	39	0·488	−0·11
12	65	61·4	1·06	24	39	0·600	1·49

a. See text, p. 19.
b. See equation (8).
* Significant at 0·05 level.

where

$$N(\mu) = \frac{1}{2\mu - 1}\binom{2\mu - 1}{\mu} \tag{4}$$

is the number of topologically distinct channel networks of magnitude μ (Shreve (1966, p. 29)). Then the approximate expected number is $8{,}724v(\mu)$. The results are shown in columns 2, 3, and 4 of Table VIII. The largest deviations from unity for the ratio of observed and expected values of N_μ are 0·86, 1·15, and 1·14 for $\mu = 3$, 8, and 11 respectively. The abundances of all networks in the magnitude range 2–5 are too low, those for 6 and 10 are about as predicted, and the others are too high. It is difficult to estimate the significance of these deviations, especially since the expected values are only approximations. Two possible methods of answering the question of significance will be explored.

One obvious solution is to make an exact calculation, explicitly taking into account the fact that the sample consists of 30 finite channel networks. The method of doing this is described below, but, as will be seen, there are some practical difficulties in carrying it out, and specific results are obtained only for magnitudes 3 and 4.

For the exact results, we need the distribution function for m, the number of subnetworks of magnitude μ in a network of magnitude n. Shreve (1966, p. 181) gives a related expression.

$$w(\mu; n) = \frac{1}{2\mu - 1}\binom{2\mu}{\mu}\binom{2(n-\mu)}{n-\mu}\bigg/\binom{2n}{n} \tag{5a}$$

where w is the probability of drawing a link of magnitude μ at random from a topologically random population of networks of magnitude n. An alternative form of this equation which is useful for hand calculations, expecially when $\mu \ll n$, is

$$w(\mu; n) = \frac{1}{2\mu - 1}\binom{n}{\mu}^2\bigg/\binom{2n}{2\mu} \tag{5b}$$

The expected value of m for given μ and n is

$$E[m; \mu, n] = (2n - 1)w(\mu; n) \tag{6}$$

To derive the distribution function for m, let $Z(m; \mu, n)$ be the number of topologically distinct channel networks of magnitude n that have exactly m subnetworks (or links) of magnitude $\mu > 1$. The possible values of m are $0, 1, 2, \ldots, m' = [n/\mu]$.

For $n < \mu$, $m' = 0$ and

$$Z(m; \mu, n) = \begin{matrix} N(n) & m = 0 \\ 0 & m > 0 \end{matrix} \tag{7a}$$

For $n = \mu$, $m' = 1$ and

$$Z(m; n, n) = \begin{matrix} 0 & m = 0 \\ N(n) & m = 1 \\ 0 & m > 1 \end{matrix} \tag{7b}$$

and for $n > \mu$, we have the general recursive relation

$$Z(m; \mu, n) = \sum_{j=1}^{n-1} \sum_{k=0}^{m} Z(k; \mu, j) Z(m - k; \mu, n - j). \tag{7c}$$

The probability density function is then $Z(m; \mu, n)/N(n)$ evaluated for all possible values of m, and the distribution function is obtained by cumulating the density function in the usual way.

As the computation of the distribution functions required considerable time on the APL system and resulted in rather bulky tables, results were obtained only for $\mu = 3$, 4 and $2 \leqslant n \leqslant 120$. The test data, chosen to conform with the limited computational results, consisted of the observed numbers of third- and fourth-magnitude subnetworks in the 11 networks with $n < 120$ and the 19 subnetworks with the largest magnitude less than 120. The quartile method is again appropriate for this kind of data and the results are shown in Table VI. As already noted in Table VIII, both third- and fourth-magnitude subnetworks are less abundant than predicted by the random model, but only for $\mu = 3$ is the deviation significant at the 0·05 level.

The exact expected values of m for magnitudes 3 and 4, calculated from equation (6) and rounded to two decimal places, are 550·95 and 345·74, compared with the approximate values of 547·1 and 342·0 used in Table VIII.

The other method for estimating the significance of the deviations of N_μ from the random model is simulation. In a small Monte Carlo experiment, 100 sets of 30 channel networks with the magnitudes given in Table III were generated by computer in a way equivalent to drawing them from topologically random populations. The total numbers of subnetworks of magnitudes 2–12, respectively, were determined for each set. The fifth column of Table VIII lists the number of simulations whose results were more extreme than the observed value of N_μ. For example, none of the 100 simulations gave a value of N_3 as low as 469, while 15 simulation values of N_9 exceeded 101. These results cannot be compared directly with the exact results for magnitudes 3 and 4, both because this test is one-tailed and because the exact results apply only for those parts of the 30 networks that have magnitudes less than 120. Also, the Monte Carlo results should be treated with some caution because of the small number of simulations, but it seems conservative to say that the random model hypothesis could be rejected at the 0·05 level for magnitudes 2, 3, 4 and 8. Again, it must be noted that these kinds of results are not independent of each other. The total number of links in the sample is fixed; if the number of second-magnitude links is less than expected, some other magnitudes must be more abundant than expected.

As mentioned previously, 1/2 is a good approximation for the expected value of E_μ/N_μ. The values listed in Table VIII exhibit an interesting pattern in that the first two are less than one-half while nine of the remaining ten are greater than 1/2 and several are considerably greater. Application of the random model to a network with 4,377 exterior links indicates that the expected number of exterior links remaining after the second- and third-magnitude subnetworks have been formed is 1,095, while in this sample there are 1,248 remaining. This observation is of course simply another aspect of the shortage of second- and third-magnitude networks. Thus, there are about 150 'excess' exterior links to be used in forming subnetworks of higher magnitude, and the run of E_μ/N_μ values greater than 1/2 uses up the major part of this excess by magnitude 10. These results are consistent with the systematic bias in diameters noted in Table V since successive combinations with exterior links will produce elongated networks.

The extreme deficiency of third-magnitude networks is now seen to result from two distinct causes. Too few second-magnitude networks are formed and, of those that do occur, too small a fraction join with exterior links. Rewriting the second part of equation (2), we have

$$N_3 = N_2(E_2/N_2)$$

and from Table VIII we see that N_2 and E_2/N_2 are about 90 and 95 per cent, respectively, of their expected values.

An estimate of the significance of the deviations of E_μ/N_μ from 1/2 can be made in the following way. If we assume that the combination processes of the subnetworks are a series of Bernouilli trials with probability 1/2 that a given subnetwork joins an exterior link, then the variable

$$z = (E_\mu - N_\mu/2)/N_\mu^{1/2}/2 \tag{8}$$

is approximately normally distributed with a mean of 0 and a variance of 1. Values of z (calculated employing the usual correction for comparing discrete variables with a distribution function for continuous variables) are given in the last column of Table VIII.

The contents of the last few paragraphs are purely formal descriptions intended to elucidate some details of the topologic structures of the networks; it is particularly important to understand that they have no implications of causality. In natural stream networks, the second- and third-magnitude subnetworks are almost certainly not formed first. The generally accepted viewpoint is that drainage networks develop by headward growth with branching from exterior links and, probably to a lesser extent, tributary formation at interior links (Dacey and Krumbein (1976)). In this case, all channels are formed as exterior links and may or may not later be converted to interior links. Interpretation of the topologic properties according to this process is difficult because the events that determine the final status of a link may occur long after it is formed. About all that can be said about this sample of networks is that not enough branching occurs in those channels which will eventually be parts of subnetworks in the magnitude range 3–12.

At this point, we may remind ourselves that it was the shortage of diameter 4 networks for $\mu = 6$ which led to the suspicion that the sample did not have enough third-magnitude networks and which in turn led to the string of calculations just completed. A deficiency for $\mu = 3$ has been confirmed but it does not appear large enough to account completely for the very small number of 3–3 pairs observed (5 as compared with 17·2 expected from Table VII). If the sample contains only 86 per cent of the magnitude 3 networks expected on the basis of the random model, then the number of 3–3 pairs should be about 75 per cent of that expected if the competing networks ($\mu = 2, 4, 5$) are themselves normally abundant. In fact, these networks also have low abundances so that the explained reduction would be less than 25 per cent and not nearly enough to account for the difference between 5 and 17·2. I have no explanation for this discrepancy. The topographic maps were reviewed for possible unconscious operator bias against 3–3 pairs but no evidence for such bias was found.

A glance at Table IIb indicates that much topologic information remains to be analyzed, e.g., path length means and standard deviations. However, as stated in the introduction, the purpose of this paper is to provide examples rather than an exhaustive treatment, and we shall close the discussion of topologic properties at this point. The most important discovery was that, in terms of the random model, the networks are slightly too elongated and that there is an associated discrepancy in the abundances of low-magnitude subnetworks.

LINK LENGTHS

Distribution functions

The basic statistical properties of the exterior and interior link lengths were reported by Smart and Werner (1976) and their results are repeated in Table IX.

Krumbein and Shreve (1970) noted that samples of link lengths for this region could be fitted reasonably well to the standard gamma density,

$$G(x; \nu, \alpha) = \frac{\alpha^{\nu} x^{\nu-1} e^{-\alpha X}}{\Gamma(\nu)} \tag{9}$$

(Here α is a scale parameter and should not be confused with the link drainage area ratio of Table IIb.) They examined two sets of data, one on fifth-magnitude networks and one on tenth-magnitude networks. In both cases, the shape parameter ν for the interior links was close to 2. For the exterior links, it was about 3·5 for $\mu = 5$ and about 6 for $\mu = 10$.

Table IX. Properties of link lengths

	Exterior	Interior
Sample size	4,377	4,347
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

The maximum likelihood estimators of ν for this data are 5·68 and 1·88 for exterior and interior links, respectively, in good agreement with the results of Krumbein and Shreve. Smart (1976, p. 114–5) observed that exterior link lengths outlined on 1:24,000 maps by the contour-crenulation method have a characteristic minimum value of 30–50 m. Thus a better fit to the data might be obtained by introducing this threshold value explicitly into the density function. For the gamma distribution this result can be achieved simply by replacing x in equation (9) by $x - x_0$, thus creating the Pearson type III density function. When this is done, the estimate of ν is reduced to 3·37 but the overall improvement in the fit is small.

In the previous section, we found it helpful to characterize the topologic properties of the networks by comparing them with the predictions of the first postulate of the random model. The analogous procedure for this section, then, is to compare the link length data with the predictions of the second postulate. Most of our analysis will be confined to the data on interior link lengths since, as stated in a previous section, that on exterior links is considered to be of lower quality. According to the second postulate, the 4,347 interior link lengths are a random sample drawn from some underlying population whose properties are determined by the local geology and climate. For test purposes, we choose the slightly more specific hypothesis that the 30 individual samples of interior link lengths came from the same population.

The raw data on link lengths consists of a set of integers ranging from 0 to 49 and having a median value of 6. Data of this type do not satisfy the assumptions employed in the standard parametric tests; consequently the nonparametric k-sample median test (Siegel (1956, p. 179)) was used to test the hypothesis. The chi-square statistic derived from the test was 108·73, compared with 42·56 for 29 degrees of freedom and the 0·05 level of significance; thus the hypothesis is emphatically rejected. This result is somewhat surprising in view of the apparent homogeneity of the environment in the region studied. One possibility is that most of the discrepancy arises from a few 'peculiar' basins. In order to check this point, the four networks with the largest contributions to the chi-square statistic (total of 59·58) were removed from the sample and the test repeated on the remaining 26 basins. The new result of 49·10 was again larger than the critical χ^2-value (37·65 for 25 degrees of freedom).

The median test was then carried out for each of the 435 different pairs of networks and the results are shown schematically in Figure 6. The characters in the 30 × 30 array indicate the magnitude of the X^2 values from the median tests. The following four ranges of X^2 were identified

$$X^2 \leqslant 0{\cdot}46$$

$$0{\cdot}46 < X^2 \leqslant 3{\cdot}84$$

$$3{\cdot}84 < X^2 \leqslant 10{\cdot}83$$

$$10{\cdot}83 < X^2$$

The values 0·46, 3·84, and 10·83 correspond respectively to $p = 0{\cdot}5$, 0·05, and 0·001 for the χ^2 distribution with one degree of freedom. The order in which the networks appear in the rows and columns has been arranged to concentrate the low values of X^2 as near the principal diagonal as possible and the high values as far away as possible.

Examination of the array suggests that the 30 networks could be divided into three or four groups in such a way that the hypothesis of the same distribution could be accepted for all or nearly all of the pairs within each group. For example, the set could be divided into three groups by choosing the first six, the next nineteen, and the last five. There is little point in trying to make an exact delineation of the groups because the choices are obviously somewhat subjective and because the use of different endpoints for the ranges would undoubtedly produce changes in the pattern. As a purely empirical observation, however, it seems acceptable to regard the link lengths as having been drawn from three or four different distributions rather than one.

The order of the networks in the array corresponds roughly to that of increasing mean interior link lengths and to that of decreasing drainage density. The most obvious guess about the clustering

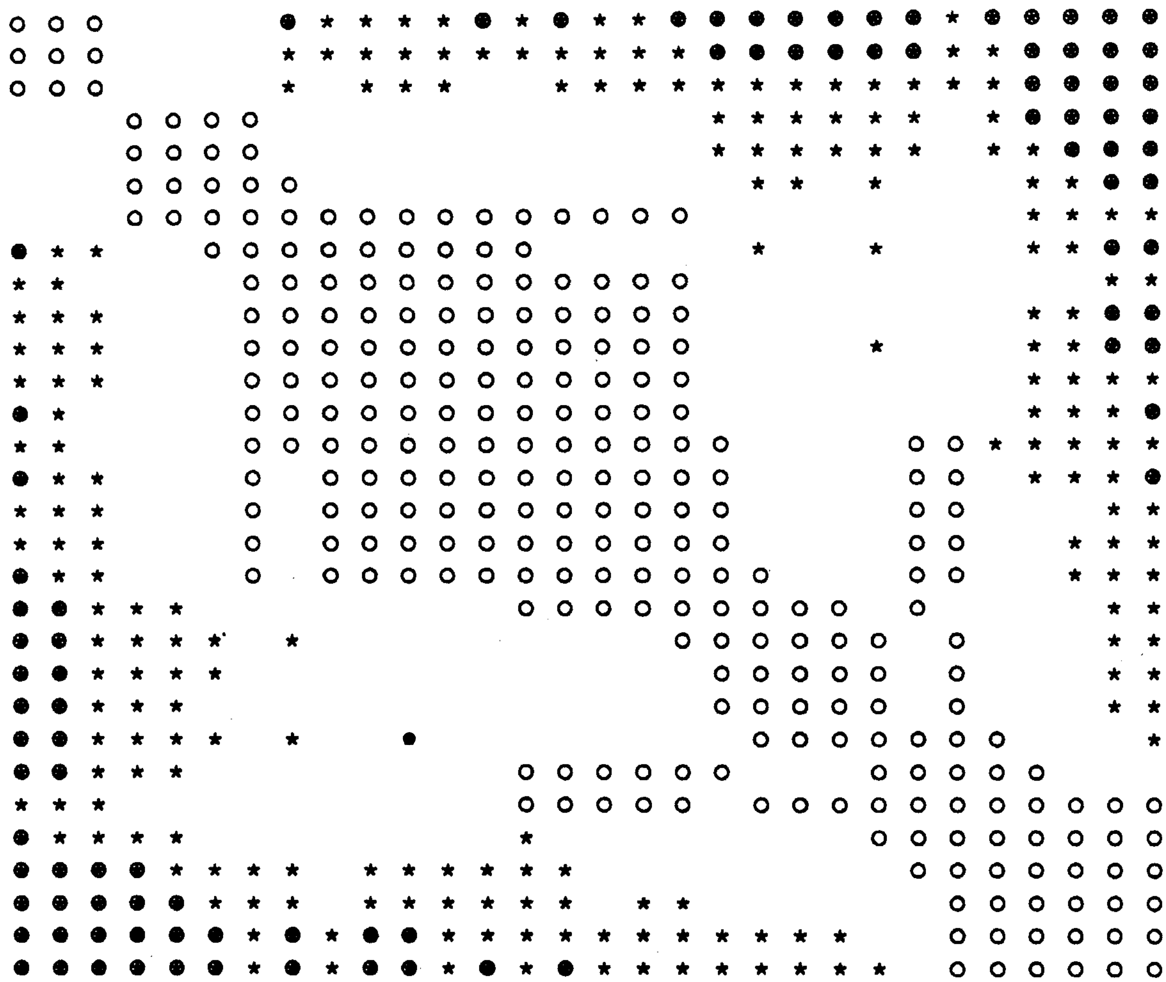


Figure 6. Pictorial representation of results of two-sample median test on link lengths for 30 networks. Symbols identify ranges of chi-square statistic X^2, as follows: open circle, $X^2 \leqslant 0{\cdot}46$; blank space, $0{\cdot}46 < X^2 \leqslant 3{\cdot}84$; asterisk, $3{\cdot}84 < X^2 \leqslant 10{\cdot}83$; asterisk and circle, $10{\cdot}83 < X^2$. Order of networks in diagram from left and top is 17, 19, 16, 15, 10, 12, 18, 11, 21, 29, 3, 1, 2, 20, 8, 24, 9, 5, 30, 7, 6, 25, 4, 27, 22, 23, 28, 14, 13, 26

effects, then, is that these properties vary somewhat over the region studied and that the members of a given group tend to be in the same immediate locality. On a small scale, there is some evidence for this speculation; pairs of networks which are either contiguous or nearest neighbours appear in sequence or with one intervening network more frequently than might be expected by chance, although no attempt has been made to fix the exact statistical significance of this pattern. On a larger scale, the members of each group are in general rather widely dispersed geographically. For example, the first group of six consists of three pairs of neighbours, but the pairs themselves come from three quite different parts of the study area. These results suggest that the spatial distribution of drainage density and related variables (such as link lengths) is patchy rather than varying uniformly with geographic position.

Magnitude dependence

Although it is clear that we must relinquish the idea that the interior link lengths were drawn from a single population, it is still possible to look for general trends in the complete set of data. One of the consequences of the second postulate of the random model is that link lengths should be independent of location in the networks. As magnitude is perhaps the most common measure of location, we shall first see if there is any evidence for a variation of link length with magnitude.

In their investigation of channel networks in the area, Krumbein and Shreve (1970) found the mean interior link length to be 0·101 km for the magnitude 5 networks and 0·087 km for the magnitude 10 networks. They considered several possible causes for the discrepancy, including a decrease of link length with magnitude, and concluded that it was largely due to systematic differences in the delineation of networks on the maps. They also noted that a sample of 485 interior link lengths collected by James and Krumbein (1969) showed no significant trend with magnitude.

The 4,347 interior link lengths in our data set were sorted by magnitude and the means and standard deviations computed for each group. The sample sizes of course fall off rapidly with increasing magnitude and for $\mu > 6$, consecutive groups were combined in such a way as to keep each sample size greater than 150. The mean link lengths for the 15 groups produced by this aggregation are plotted as a function of magnitude in Figure 7 (curve I). The ordinate axis is scaled in the actual units of measurement (15·24 m). The plot suggests a slight increase of mean link length with magnitude but as the standard deviations are typically about 5·0, the possible significance of the trend is not immediately obvious.

The two-sample median test was applied by first dividing the complete sample of interior link lengths into a low magnitude group ($\mu \leqslant 6$) and a high magnitude group ($\mu > 6$) of approximately equal size and then dividing these groups according to whether the lengths do or do not exceed the combined median. The X^2 value was 12·01 while the critical value for the 0·05 level and one degree of freedom is 3·84. The low magnitude group had too few lengths larger than the median while the high magnitude group had too many. Thus there is a definite tendency for link lengths to increase with magnitude.

Despite the highly significant X^2 value, this effect quickly becomes diminished when tested with larger resolution of the data groups, no doubt because of the heterogeneous quality of the individual network link length distributions. For example, application of the k-sample median test to the 15 groups described above allows acceptance of the hypothesis that all groups are from populations with the same median. There is, however, a strong clue to the two-sample result in the pattern of deviations; all five of the groups with $\mu \leqslant 6$ have fewer lengths exceeding the median than expected while nine of the ten groups with $\mu > 6$ have more.

In a similar manner, the results obtained by applying the two-sample median test with low and high magnitude groups to the data for individual networks is not very conclusive. Five of the 30 results were significant at the 0·05 level; of these five, the signs of the deviations indicated an increase of link length with μ in four cases and a decrease of link with μ in one case. For all 30 networks, regardless of the size of X^2 an increase of link length with μ was indicated in 22 cases and a decrease in eight cases. Thus results obtained with only one or two networks can be completely at variance with those obtained from a large sample. This point may explain the diverse results obtained by Smart (1969) in making tests of the relation between $\bar{l}_i$ and μ on ten networks chosen from ten different environments.

A new interior link classification scheme

In the topologic analysis, it proved useful to classify subnetworks according to whether they combined with an exterior or an interior link at their downstream ends. A similar classification can obviously be applied to links and it seems worthwhile to see whether the two kinds, which we shall designate as *IE* and *II*, have any differences in their length properties. Mock (1971) proposed an analogous classification for exterior links but he did not investigate the length distributions. Abrahams and Campbell (1976), however, recently found that the two groups have significantly different length distributions, with those exterior links that join another exterior link having more short and fewer long links. Mock designated the two groups as *S* and *TS* according to whether they joined an exterior or interior link, but we shall use the terms *EE* and *EI* to conform with the interior link notation.

Table X. Link length data

Type	N	$\bar{l}$ (km)	σ (km)	C.V.	Max. (km)	Min. (km)
IE	2,291	0·089	0·065	0·731	0·472	0
II	2,056	0·137	0·090	0·657	0·747	0
EE	2,086	0·130	0·070	0·543	0·747	0·046
EI	2,291	0·185	0·090	0·495	0·945	0·030

The link numbers are related to the subnetwork frequencies E_μ and I_μ. If N_{AB} is the total number of links of type AB in the sample, then

$$\begin{aligned} N_{EE} &= E_1 \\ N_{EI} &= I_1 \\ N_{IE} &= \sum_{\mu=2}^{267} E_\mu \\ N_{II} &= \sum_{\mu=2}^{267} I_\mu \end{aligned} \tag{10}$$

Outlet links are arbitrarily designated *II*.

The 4,377 exterior links and 4,347 interior links were sorted into their respective categories and the mean lengths were computed. The results are shown in Table X. In both cases there is a surprising difference in the two groups with the *XI* links being about 50 per cent longer than the *XE* links. The results on the relative lengths of *EE* and *EI* links are qualitatively consistent with those of Abrahams and Campbell (1976). However, another feature of their data, an increase in *EI* link length with magnitude of link joined, does not occur in ours.

The two groups of interior links were sorted into the same magnitude categories used for all interior links and the length distributions are shown in Figure 7. This graph indicates quite clearly that the results for the complete sample of interior links are simply some sort of an average for two quite different components. It is interesting that the classification method, which is purely topologic, displays

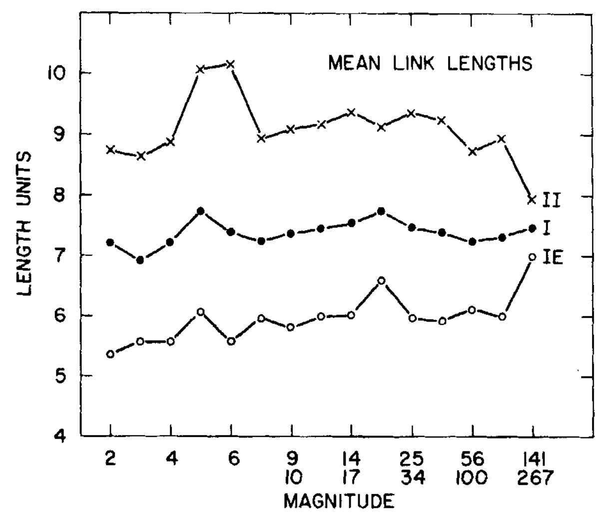


Figure 7. Mean interior link length versus magnitude. *I* indicates all interior links of that magnitude and *II* and *IE* the subclasses described in the text. Double figures on ordinate indicate a range of magnitudes. Missing magnitude ranges can be deduced from those on each side. 1 length unit = 15·24 m

such a strong differentiation in a geometric property. This discovery also raises the possibility that the two groups of links may be regarded as independent random variables with a common distribution for each, thus generalizing the second postulate of the random model. Application of the *k*-sample median test, however, convincingly rejects this hypothesis in both cases. In general, the sizes of *II* and *IE* lengths tend to rise and fall together; the 30 values of $\bar{l}_{II}/\bar{l}_{IE}$ lie in the fairly narrow range 1·23–1·84.

In Figure 7, the *IE* mean lengths show a rather steady increase with magnitude, while the *II* behaviour is more erratic with a suggestion of a slight downward trend. The two sets of data were each divided into a low magnitude group and a high magnitude group, as was previously done for all interior link lengths together. The median test was again applied to the null hypothesis of no change in median with magnitude. The hypothesis was rejected, with χ^2 statistics of 18·36 and 12·91 for the *IE* and *II* groups respectively. In both cases, the signs of the deviations indicated larger link lengths at high magnitudes. This result for the *II* links may seem inconsistent with Figure 7 and indeed the Spearman rank correlation coefficient between $\bar{l}_{IE}$ and μ is $-0{\cdot}2$. The reason for this apparent discrepancy is that the low magnitude ($\mu \leqslant 6$) groups have a relatively large number of very long links (500–750 m) which elevate the means without affecting the medians. The median values are plotted for comparison in Figure 9. The apparent convergence of both means and medians of *II* and *IE* links for the largest magnitude category is puzzling. It may be that the distinction between the two groups gradually disappears with increasing magnitude but more data at higher magnitudes would be required to test this possibility.

Cis-trans links

James and Krumbein (1969) proposed a classification of interior links of main channels as cis and trans links, according to whether the tributaries at the upstream and downstream ends joined from the same side or from opposite sides. In a sample of 485 links taken from this Kentucky area (Middle Fork of Rockcastle Creek, Thomas and Inez quadrangles), they found 192 cis links and 293 trans links. The random model of course predicts that the two kinds of links should occur with equal frequency, while the probability of drawing only 192 cis links in a sample of 485 if the underlying population did contain equal proportions is less than 10^{-5}. This result gave one of the first clear indications of the limitations of the random model in describing the topologic structure of channel networks.

Our set of 30 networks contained 880 mainstream links of which 337 or 38·3 per cent were cis, in close agreement with the James and Krumbein value (39·6 per cent). The two-way contingency table was used to test for a possible relation between the cis-trans classification and the *II–IE* classification.

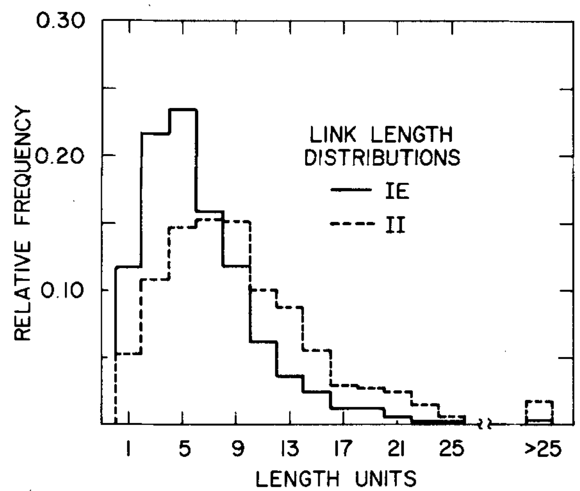


Figure 8. Histograms of link lengths for *II* (solid line) and *IE* (dotted line) classes. 1 length unit = 15·24 m

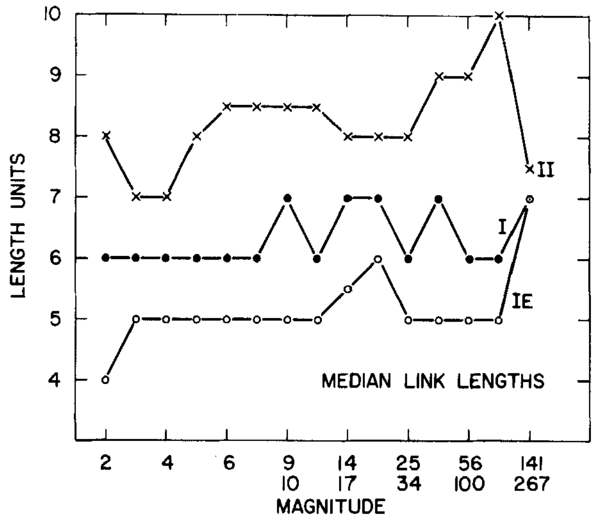


Figure 9. Median link lengths versus magnitude. I indicates all interior links of that magnitude and II and IE the subclasses defined in the text. Double figures on ordinate indicate a range of magnitudes. Missing magnitude ranges can be deduced from those on each side. 1 length unit = 15·24 m

The χ^2 statistic was 0·40, compared with 3·84 for the 0·05 level of confidence. Thus there is no reason for rejecting the hypothesis that the two methods are independent.

In conclusion, the picture of the geometric structure of channel networks which we have developed is rather more complex than the idealized version of the random model. Those factors which determine the scale of the network structure, such as drainage density and mean link lengths, show an unexpectedly large, and perhaps patchy, variation over an apparently uniform region. Also, for each individual network, regardless of scale, there are two distinct kinds of interior links, with the II type being appreciably longer than the IE type. The exterior links have received less study, but there is definite evidence of two types here also, with the ratio $\bar{l}_{EI}/\bar{l}_{EE}$ ranging between 1·10 and 1·99. The interior link lengths tend to increase with increasing magnitude, with the effect being more pronounced and better defined in the IE group.

MISCELLANEOUS RESULTS

One point of considerable interest about drainage basin composition is the way in which various properties change as the size of the basin increases. Although the present sample of data is not the optimum design for displaying such effects, some observations and inferences can be made. There are three possible ways of studying size effects with this data, each with its own advantages and disadvantages. The most obvious way is simply to correlate the watershed vectors against the area vector or the network magnitude vector. The main problem with this procedure is that the rather large variation from network to network for such properties as drainage density and mean link length combined with the rather small sample size may result in some spurious correlations.

Another possibility is to divide the individual networks into nested subnetworks and to study the relations between the properties of the subnetworks and their respective areas of magnitudes. This method presumably confines the data to a fairly uniform region but it has the disadvantage that the observations are not independent so that it is difficult to assign a quantitative measure of significance to the results. Finally, individual basins can be divided into non-overlapping areas such as those defined by selecting the nested subnetworks. This produces independent observations but the range of sizes is greatly reduced. Also, the channel nets in some sections are complete channel networks while in others they are fragments of channel networks.

All three procedures were employed to a limited extent in this study. For the second and third methods, three networks, numbers 26, 3, and 30 of Table III, were chosen because they have, respectively, low, medium, and high drainage density. The basin area was apportioned by first choosing the subbasins defined by the major tributaries to the mainstream and then making further subdivisions so as to create sections containing 15 or 20 links. An example is shown in Figure 10 for basin number 26, Canoe Creek above Little Fork, Canoe quadrangle.

The three closely related parameters D, F and K have all been used as measures of the degree to which the basin area is dissected by channels. Their exact relationship is given by

$$K = D\bar{l} = F\bar{l}^2 \tag{11}$$

where $\bar{l}$ is the mean link length including both exterior and interior links. A considerable part of the geometric composition of channel networks is determined by the way in which these variables change with basin size and by their interrelations. (It should be noted that all four quantities are functions of the more basic and highly correlated variables A, L, and $2n - 1$, which are themselves measures of basin size.) Some results are given in Table XI; the Spearman rank correlation coefficient is used to avoid having to transform the data to produce quasi-linear relationships.

For the watershed data, only the mean link length $\bar{l}$ is significantly (0·05 level) correlated with area. The positive coefficient is of course consistent with the results previously obtained for the variation of interior link length with magnitude. Both D and F have weak negative correlations with area; as these derived quantities are both ratios with A in the denominator, a negative correlation is to be expected unless the numerators, L and $2n - 1$, increase in at least a linear fashion with area. Of the six pairs of interactions between the four variables, D, F, K and $\bar{l}$ only that between F and K is not significant.

The most notable result for the three networks is that numbers 3 and 30 behave in roughly the same way and quite differently from number 26. In particular, D and $\bar{l}$ have large negative correlations for the two networks but are only weakly correlated for Canoe Creek. Because of the close relationship indicated by equation (11), this difference must be reflected elsewhere in the table, and we see that

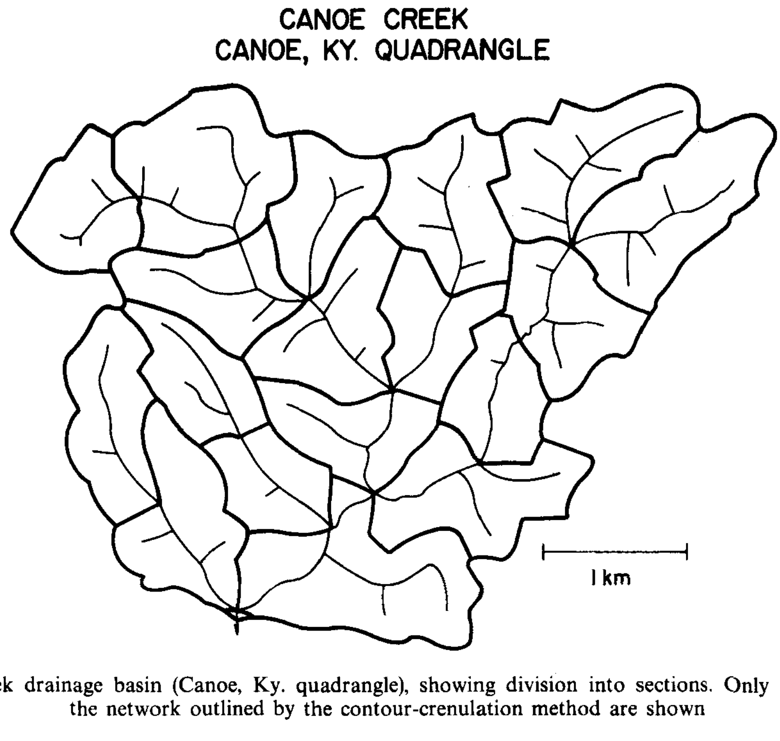


Figure 10. Canoe Creek drainage basin (Canoe, Ky. quadrangle), showing division into sections. Only the major channels of the network outlined by the contour-crenulation method are shown

Table XI. Spearman rank correlation coefficients

Vars.		All networks	Nested data 26	3	30	Non-nested data 26	3	30
D	A	−0·317	0·191	−0·373	−0·193	−0·608*	−0·627*	−0·784*
F	A	−0·330	−0·035	−0·391	−0·346	−0·579*	−0·643*	−0·872*
K	A	−0·026	0·323	−0·114	0·027	−0·146	−0·333	−0·498*
$\bar{l}$	A	0·441*	0·205	0·298	0·392	0·387	0·477*	0·737*
λ	A	−0·040	−0·249	−0·521	−0·414			
α	A	−0·043						
β_T	A	0·284	0·457	0·046	0·596			
β_G	A	0·204	0·565	0·046	0·517			
D	F	0·888*	0·672	0·958	0·945	0·709*	0·905*	0·925*
D	K	0·451*	0·205	0·470	0·779	0·670*	0·666*	0·815*
D	$\bar{l}$	−0·532*	−0·251	−0·712	−0·775	−0·084	−0·545*	−0·639*
F	K	0·078	−0·477	0·313	0·632	0·036	0·329	0·590*
F	$\bar{l}$	−0·799*	−0·835	−0·856	−0·895	−0·678*	−0·835*	−0·858*
K	$\bar{l}$	0·412*	0·816	0·123	0·344	0·635*	0·213	−0·133
Sample size		30				18	23	23
r_s (0·05)		0·364				0·475	0·419	0·419

* Significant at 0·05 level.

K and $\bar{l}$ have a strong positive correlation for Canoe Creek and only weak correlations for the other two networks. Similar remarks apply to F and K.

In general, the pattern of correlation coefficients for the non-nested data is surprisingly similar to that for the nested data. The principal difference is perhaps the large negative correlation of D and K with A. It is of course not all unreasonable that this effect should be stronger for the small areas (generally less than 1 km^2) involved in the non-nested data.

One of the first things noticed in the link-based approach to drainage basin composition was that the properties associated with exterior and interior links, such as length and drainage area, differed significantly. With the measurements on A_e and A_i provided in this paper, the variables K, D, and F can be separated into exterior and interior components in the same way. Of course, in view of the discovery of the previous section that there are two kinds of exterior links and two kinds of interior links, a division into four components would be still more appropriate. The requisite area data is not available for this step; however, some worthwhile information can be obtained from the conventional separation into exterior and interior parts. The variables K_e, K_i, D_e, D_i, F_e, and F_i are defined by equations completely analogous to those for their total basin counterparts. In addition

$$\frac{F_e}{F_i} = \frac{1}{\alpha}, \qquad \frac{D_e}{D_i} = \frac{\lambda}{\alpha}, \qquad \frac{K_e}{K_i} = \frac{\lambda^2}{\alpha} \tag{12}$$

Descriptive statistics for the nine variables are given in Table XII. The differences between F_e and F_i and between K_e and K_i are clearly highly significant. The mean values of D_e and D_i on the other

Table XII. K, D, and F: exterior and interior components

Property	K_e	K_i	K_e/K_i	D_e (km^{-1})	D_i (km^{-1})	D_e/D_i	F_e (km^{-2})	F_i (km^{-2})	F_e/F_i
Mean	0·949	0·614	1.56	6·07	5·57	1·09	39·4	51·9	0·778
S.D.	0·171	0·061	0·33	1·08	0·62	0·17	10·0	13·3	0·167
C.V.	0·181	0·099	0·212	0·178	0·112	0·154	0·254	0·257	0·215
Max.	1·322	0·719	2·42	8·96	7·08	1·47	66·4	85·1	1·256
Min.	0·583	0·511	0·91	4·32	4·56	0·77	24·0	29·4	0·504

Table XIII. Topologic and geometric ratios

Property	$d/\bar{j}_e$	$L'/\bar{J}_e$	d/s_e	L/S_E	$\bar{j}_e/s_e$	$\bar{J}_E/S_E$
Mean	1·70	1·69	4·20	4·33	2·48	2·59
S.D.	0·11	0·14	0·38	0·39	0·32	0·36
C.V.	0·07	0·08	0·090	0·090	0·13	0·14
Max.	1·86	1·99	5·00	5·32	3·33	3·48
Min.	1·47	1·39	3·54	3·77	1·90	2·05
$r_S(X, A)$	0·571	0·409	0·016	0·301	0·345	0·462

hand, are not greatly different, although the Kolmogorov-Smirnov test just barely rejects (0·05 level) the hypothesis that the two sets of data have the same distribution. Thus the Kentucky drainage networks are formed with approximately equal interior and exterior drainage densities, but with $K_e > K_i$ and $F_e < F_i$. The two inequalities are of course not independent because of the relation $D = (KF)^{1/2}$.

As mentioned in the beginning, one very important feature of drainage basin composition is the close relation between topologic and geometric properties (Shreve (1975)). This point did not receive much specific attention in this paper because it has been covered in some detail by Smart and Werner (1976), but one additional example is perhaps appropriate. Table XIII gives descriptive statistics for the ratios of the maximum, mean, and standard deviation of the exterior path lengths and for the corresponding geometric ratios. The ratios are all quite stable, with low coefficients of variation, but more important is the excellent agreement between all corresponding statistics for each topologic–geometric pair. This result demonstrates again the very practical possibility of using topologic results, which are generally easy to measure, to obtain good estimates of geometric results, which are frequently difficult to measure. Despite the small coefficients of variation, the variables do seem to show a significant trend with basin size. The last row in the tables gives the Spearman rank correlation coefficients with area (0·364 = critical value for 0·05 level). The values of r_s agree well for two of the pairs but not for d/s_e and L'/S_E. The ratios listed in the table are thus appropriate only for the ranges of magnitudes and areas encountered in this sample; one may speculate that for a random model population the ratios would approach 2, 4, and 2 respectively, as the basin size becomes larger and larger.

THE SLOPE-CRITERION METHOD

Although the Strahler contour-crenulation method is the most common way of outlining channel networks on 1:24,000 maps, it has been subjected to considerable criticism. One objection concerns the general accuracy of the maps; as examples, Coates (1958, p. 24–25) describes a case in which the maps provide insufficient detail and Maxwell (1960 p. 13, 60) a case in which there are major errors in the contours. Another objection is that the somewhat subjective nature of the judgments required to identify the smallest crenulations introduces an unacceptable degree of operator variation. Finally, there is some doubt as to whether exterior links and sources identified in this way really correspond to the actual exterior links and sources, even for maps of proven accuracy.

As far as the Kentucky networks are concerned, the question of operator variation seems reasonably well settled, by the Krumbein–Shreve and Smart–Werner (1976) experiments described in a previous section. Perhaps more important, however, is the good agreement between this paper and Krumbein and Shreve (1970) and James and Krumbein (1969) on some rather subtle details of the composition of the networks, such as the excess of large diameters and the cis–trans ratio.

Field studies are of course required to provide definite answers to the other two criticisms and none have been carried out for the Kentucky area. In connection with the identification of sources, Krumbein and Shreve (1970) have argued that, given accurate maps, the contour-defined sources have a geomorphological significance of their own, regardless of how well they correspond to natural sources, and are thus a suitable basis for defining channel networks.

Shreve (1974, p. 1172) has proposed another method of delineating channel networks which greatly reduces, if not eliminates, the problems discussed above. The source of a channel is identified as the point where it is crossed by the highest contour spaced more than some predefined critical distance from the next lowest contour. This procedure will be called here the *slope-criterion* (SC) method to distinguish it from the Strahler contour-crenulation (CC) method. For the Kentucky maps, which have contour intervals of 12·2 m (40 ft), Shreve selected 61·0 m (200 ft) as the critical distance. This choice has the general (and intentional) effect of placing the sources at heads of valleys, rather than along their sides, and of reducing their number to about one-fifth that obtained by the CC method. The difference in the results of the two methods is illustrated by Figure 11 for network number 8, Blacklog Fork, Kermit quadrangle.

In addition to its other advantages, the SC method obviously requires considerably less labor than the CC method. It is therefore of some interest to make a detailed comparison of the results obtained by the two different procedures. It is clear that some properties (such as mainstream length) will be very little changed, and some (such as network magnitude) will be changed by an approximately-constant factor, while for still others the end result is difficult to estimate. One would also like to know whether the correlations between various properties are changed significantly and whether certain special results of the CC analyses, such as the diameter distributions and the cis–trans ratio, persist in the SC method.

The channel networks for the 30 basins in our sample were outlined by the SC method, using Shreve's value for the critical slope. Link lengths were measured to the nearest half mm (12 m full scale). It

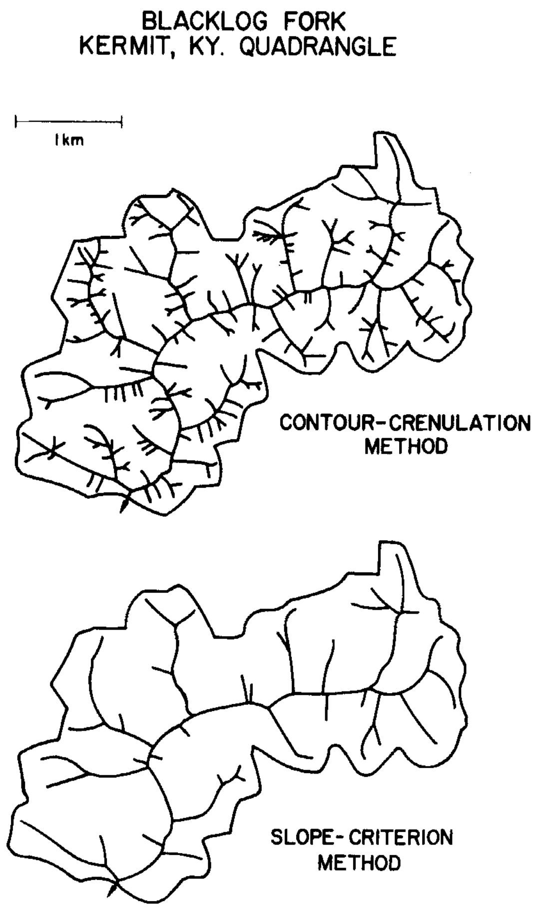


Figure 11. Comparison of channel networks for Blacklog Fork drainage basin (Kermit, Ky. quadrangle) as outlined by contour-crenulation and slope-criterion methods

turned out that three of the networks (numbers 11, 12, and 14) were subnetworks of the Middle Fork of Rockcastle Creek above Inez, which had previously been analyzed by Shreve (1974). Professor Shreve kindly loaned me his data, thus making possible a quantitative check on operator variation in the SC method. For the three networks, my results for the total number of sources, mean exterior link length, and mean interior link length were 80, 0·269 km, and 0·312 km, while the corresponding values obtained by Shreve were 78, 0·265 km, and 0·296 km. Seventy-five sources were identified by both Smart and Shreve, five by Smart only, and three by Shreve only. At first sight, this result is rather discouraging since the relative differences are about the same as in the CC method. However, a detailed examination of the maps showed that all differences of opinion about source identification occurred for one special situation. Channel segments that have a small flood plain 200–300 m across are occasionally accompanied by a single contour crenulation perpendicular to the channel which may or may not be interpreted as a short exterior-link tributary. When an objective (but somewhat arbitrary) rule was devised to take care of this situation, the two sets of sources were brought into complete agreement. Also, the accompanying readjustment of link lengths and classifications, brought the values of mean interior link length to within about 0·007 km of each other. The only discrepancy remaining was an unexplained difference of about 2 per cent in total channel length.

The SC networks range in magnitude from 7 to 69 and the total number of sources is 863. Tables XIV and XV give some data analogous to that presented in Tables III and IV for the CC analysis.

To help compare the results obtained by the two methods, Table XVI shows some descriptive statistics

Table XIV. Properties of individual drainage basins—SC method

Network No.	n	d	p_e	A (km^2)	L (km)	D (km^{-1})	L' (km)	$\bar{l}_e$ (km)	$\bar{l}_i$ (km)
1	57	27	875	12·29	33·3	2·71	6·68	0·328	0·262
2	33	21	405	9·27	20·9	2·26	7·06	0·323	0·320
3	38	16	365	10·60	23·3	2·20	4·14	0·323	0·297
4	69	26	1055	12·62	35·4	2·80	6·07	0·284	0·232
5	35	13	322	9·28	19·2	2·07	4·42	0·277	0·278
6	35	21	478	9·92	21·3	2·15	6·22	0·312	0·306
7	45	28	659	10·29	25·8	2·50	7·15	0·307	0·272
8	35	22	455	8·02	20·7	2·59	5·06	0·337	0·263
9	26	18	299	6·47	12·8	1·98	4·15	0·281	0·219
10	24	15	237	5·49	12·6	2·30	3·84	0·286	0·250
11	24	17	248	6·08	13·2	2·18	5·23	0·251	0·313
12	27	14	267	6·23	15·7	2·53	3·89	0·274	0·321
13	34	19	389	7·42	18·0	2·42	4·94	0·274	0·263
14	27	11	211	5·93	15·1	2·55	3·24	0·264	0·307
15	18	10	118	5·64	10·2	1·81	2·80	0·341	0·241
16	11	9	66	3·56	6·1	1·73	3·16	0·273	0·314
17	14	8	81	3·47	8·4	2·41	3·06	0·247	0·379
18	11	8	60	1.83	4·7	2·58	1·56	0·243	0·205
19	7	7	34	2·23	4·8	2·14	2·77	0·367	0·368
20	10	9	58	2·02	5·5	2·73	2·35	0·308	0·271
21	27	12	225	5·77	14·1	2·45	3·22	0·269	0·264
22	11	6	51	1·92	5·0	2·61	1·93	0·187	0·298
23	20	11	144	4·31	10·6	2·45	3·31	0·287	0·254
24	17	12	139	4·88	10·6	2·17	3·86	0·296	0·349
25	37	21	508	7·47	20.9	2·80	5·50	0·307	0·265
26	60	19	710	13·68	33·6	2·45	5·59	0·270	0·295
27	28	14	251	7·34	17·9	2·44	3·89	0·352	0·299
28	16	12	118	3·71	10·0	2·68	3·62	0·374	0·265
29	31	22	376	9·68	22·2	2·30	6·42	0·343	0·387
30	36	25	536	10·17	21·8	2·14	5·46	0·302	0·311

Table XV. Descriptive statistics of some geomorphic properties—SC method

	$D\,(\mathrm{km}^{-1})$	$F\,(\mathrm{km}^{-2})$	K	λ	α
Mean	2·37	8·22	0·692	1·05	1·64
S.D.	0·27	1·43	0·095	0·19	0·49
C.V.	0·115	0·174	0·138	0·186	0·297
Max.	2·80	11·44	0·862	1·42	3·47
Min.	1.73	5·82	0·492	0·63	0·73

on ratios of the type X_c/X_s where X is a geomorphic variable and the subscripts C and S refer to measurements by the contour-crenulation and slope-criterion techniques, respectively.

We first check to see if there are any variables whose SC value is sufficiently close to the CC value that it can be used as an estimate. If we arbitrarily define 'sufficiently close' as meaning that the ratio mean is within 10 per cent of unity and the coefficient of variation less than 0·2, then there are four: L', $\bar{J}_E$, S_E and β_G. The mainstream length was previously identified as a variable that would be relatively insensitive to the change in methods, and of course, the shape parameter β_G is a function only of L' and A. It is, however, rather surprising that $\bar{J}_E$ and S_E should remain so constant considering the drastic change in network properties illustrated in Figure 11. Something of the same behaviour can be noted in the topologic counterparts d, $\bar{j}_e$, and s_e but here the proportionality factor is about 2·6 and the coefficients of variation are much larger.

It would be highly desirable to have a theoretical analysis of the effect that changing the criterion for source identification has on the drainage basin parameters. Although a general solution of this problem seems out of the question, there is one special case for which approximate results can be obtained very easily. This case occurs when the method which produces the network of lower magnitude

Table XVI. Descriptive statistics for X_c/X_s

X	Mean	S.D.	C.V.	Max.	Min	Estimate of X_c/X_s Algebraic	Numerical
n	5·51	1·85	0·335	11·29	3·19	$\rho = n_c/n_s$	4
L	2·51	0·55	0·218	4·12	1·71	$1 + \lambda_c$	2·4
D	2·51	0·55	0·218	4·12	1·71	$1 + \lambda_c$	2·4
F	5·63	1·96	0·348	12·08	3·20	ρ	4
K	1·15	0·24	0·209	1·81	0·85	$(1 + \lambda_c)^2/\rho$	1·4
$\bar{l}_e$	0·54	0·10	0·181	0·76	0·34	$\lambda_c/2$	0·7
$\bar{l}_i$	0·40	0·09	0·221	0·56	0·21	$2/\rho$	0·5
λ	1·43	0·41	0·288	2·77	0·91	λ_c	1·4
$\bar{a}_e$	0·20	0·06	0·306	0·32	0·08	$\alpha_c/2(1 + 2\alpha_c k)$*	0·29–0·18*
$\bar{a}_i$	0·19	0·06	0·326	0·30	0·08	$1/2(1 + 2\alpha_c(1 - k))$	0·21–0·50
α	0·88	0·30	0·346	1·83	0·45		1·30–0·36
L'	1·04	0·04	0·040	1·13	0·98	1	1
P_E	5·27	1·75	0·333	11·34	2·93	$2\rho^{1/2}$	4
$\bar{J}_E$	0·96	0·08	0·082	1·21	0·81	$2\rho^{-1/2}$	1
S_E	1·07	0·19	0·172	1·93	0·86		1
β_G	1·08	0·09	0·081	1·27	0·95	1	1
d	2·68	0·76	0·282	5·14	1·95	$\rho/2$	2
p_0	14·40	9·36	0·650	51·77	5·53	$\rho^{3/2}$	8
$\bar{j}_e$	2·44	0·61	0·248	4·59	1·74	$\rho^{1/2}$	2
s_e	2·75	1·03	0·376	6·23	1·81		2
β_T	1·33	0·45	0·341	2·66	0·64	$\rho/4$	1

*The parameter k is the fraction of CC exterior basin area that drains into SC exterior links. The range of computed values corresponds to $0·5 \leqslant k \leqslant 1$.

simply removes all of the exterior links, and nothing else, from the larger network. As the CC and SC networks afford a rather rough approximation to this idealized model (863 SC sources compared to 1,043 obtained by removing the CC exterior links), some results are listed in Table XVI for comparison. The algebraic expressions were obtained by employing results of the random model and by neglecting terms of unity compared to n_c and n_s. The ratio $\rho = n_c/n_s$ is given explicitly to show where it appears in the calculations, but since the expressions apply only to the random model it would not be appropriate to assign it any value other than 4. The mean values of λ_c and α_c from Table IV (1·43 and 1·34) were used in obtaining the numerical results. Another parameter outside the random model that was needed to estimate the A_e, A_i, and α ratios is the fraction of CC exterior drainage area that is drained by the SC exterior links. Numerical values were computed assuming that this ratio, k, lies in the range 0·5–1. One interesting feature of the model is that $\lambda_s = 1$, the reason being that if the *EI* links are randomly located in the CC networks, removing them produces a new set of links, all with the same length distribution regardless of whether they are identified as exterior or interior in the SC networks. The observed mean of λ_s is in fact 1·05, but this agreement is much better than could be expected, considering the disparity between the model and the actual system. It is clear from inspection of the entire table, however, that the model does satisfactorily reproduce the general behaviour of the X_c/X_s ratios.

Other aspects of the relation between the CC and SC methods will be developed as the results of the SC composition analysis are obtained.

ANALYSIS OF THE SC RESULTS

The analysis of the SC networks will necessarily be considerably less detailed than that for the CC case because of the fewer number of links involved. Wherever possible, however, analogous investigations will be carried out for both cases to provide a basis for comparison.

Topologic properties

Table XVII shows the observed and theoretical diameter distributions for magnitudes 4 through 7. The random model hypothesis is easily accepted for all four sets of observations. There is some suggestion of a tendency towards too many large diameter networks but the data are too skimpy to properly evaluate its significance. Application of the quartile test to the observed diameters of the 30 complete networks give occupation numbers of 5·82, 5·31, 3·59, and 15·28 for the quartiles listed in order of increasing d. The fit is obviously quite poor since over half of the total falls in the largest diameter cell; we have $X^2 = 11{\cdot}12$ compared to 7·81 for the 0·05 level.

Values of N_μ and E_μ for μ in the range of 2–12 were determined and compared with the random model. There were only two instances of serious disagreement. The observed number of second-magni-

Table XVII. Theoretical and observed diameter distributions—SC case

μ	4		5		6		7	
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
d								
3	16	14·2						
4	55	56·8	16	21·4	5	5·42	7	10.33
5			34	28·6	16	18·10		
6					17	14·48	13	13·15
7							11	7·52
	71[a]	0·29[b]	50	2·41	38	0.72	31	2·69
		3·84[c]		3·84		5·99		5·99

a. Sample size.
b. Chi-square statistic, X^2.
c. Critical value for 0·05 level of significance.

tude networks was 198, which is considerably less than the theoretical expected value of 219·78. To obtain a quantitative estimate of the significance of this discrepancy, 100 simulations of the kind described in the CC discussion were carried out using the SC network magnitudes. The average of the 100 results for N_2 was 219·89 and the range was 203–239. Thus none of the simulation values was as low as 198. The other notable discrepancy was that 16 of 18 ninth-magnitude links joined an exterior link downstream; there seems no reason to attribute the result to anything other than an isolated fluctuation. Otherwise, the observed values of E_μ and N_μ conformed fairly closely to the random model predictions.

A tabulation of the cis–trans links in the mainstreams showed that 96 of the total of 215 (44·7 per cent) were cis links. The hypothesis that the sample was drawn from a population containing equal frequencies can be accepted at the 0·05 level.

There is a considerable degree of similarity between the topologic results for CC and SC networks. In both cases, the diameters display a small but systematic bias towards large values. Also, for both cases there are far too few second-magnitude links. On the other hand, only the CC networks show significant discrepancies in abundance for other low magnitude links. Both cis–trans ratios are less than one but the difference is significant for CC networks and not for SC networks. However, it may be noted that the standard test for the difference of two proportions (Hoel (1962, p. 148–151)) does not reject at the 0·05 level the hypothesis that the two sets of data were drawn from the same binomial population. The failure to reject either of two distinct and incompatible hypotheses (that the SC cis–trans ratio is one, and that it is equal to the CC ratio) simply reflects the fact that the tests are designed to avoid rejecting a correct hypothesis. To guard against accepting a false hypothesis (type II error) the confidence level must be raised. The critical probabilities derived from the test statistics in the two cases are 0·12 and, 0·08 respectively.

Link lengths

In the SC method the exterior and interior link length measurements are considered to be equally accurate and reproducible, so the two groups will be treated on the same basis in the analysis. The first step was to use the k-sample median test to compare the length. X^2 was 34·45 for exterior links and 22·26 for interior links, both values being appreciably lower than the 0·05 critical value of 42·56. Thus the agreement with the random model is much better than in the CC case, although, as can be determined from the data of Tables III and XIV, the corresponding coefficients of variation of link length means are almost exactly the same (0·134 and 0·131 for exterior links, 0·151 and 0·157 for interior links). The point is that this degree of variation is consistent with a single population for samples of SC size but not for those of CC size. The median test is of course not particularly strong and it is quite conceivable that more detailed tests might display significant differences between the distributions for individual networks; more data is needed, however, before such tests could be profitably carried out. In any event, it is clear that the SC method produces link length samples that are much more homogeneous than those of the CC method.

In order to test for possible dependence of link lengths on magnitude, the interior link sample was divided into a low magnitude group and a high magnitude group as described in the CC analysis. The two-sample median test showed no significant difference in the medians of the two groups.

The results of the k-sample median test certainly do not exclude the possibility that both exterior and interior links are divided in two groups according to their downstream junctions, as in the CC case. Some statistical properties of the link lengths for the four classes are given in Table XVIII. The

Table XVIII. Link length data—SC method

Type	N	$\bar{l}$ (km)	S.D. (km)	C.V.	Max. (km)	Min. (km)
IE	467	0·248	0·173	0·697	1·608	0·012
II	366	0·329	0·197	0·597	1·188	0·012
EE	396	0·283	0·204	0·720	1·164	0·060
EI	467	0·311	0·195	0·626	1·392	0·072

ratios of the means $\bar{l}_{II}/\bar{l}_{IE}$ and $\bar{l}_{EI}/\bar{l}_{EE}$ are 1·3 and 1·10, respectively. It is apparent from the table that the *IE* length distribution is different from the other three, and application of the Kolmogorov–Smirnov test shows that the hypothesis of identical distributions can be rejected at the 0·05 level for the other three pairs as well. It is particularly interesting that the distinction between *XE* and *XI* links ($X = E, I$) noted for the CC method also occurs in the SC results.

Miscellaneous results

Not much can be said about the dependence of the SC variables on area. As in the CC case, seven of the eight watershed properties listed in Table XI have values of r_s below the 0·05 critical level. The analysis of area effects by dividing the individual basins into sub-areas is not practicable for this set of SC networks because of the small number of links per sub-area.

Figure 12 is a pictorial display of the intercorrelations among a set of 16 selected watershed variables. The symbols indicate values of the Spearman rank correlation coefficient for individual pairs. The range divisions 0·306 and 0·465 correspond to the 0·05 and 0·01 levels of significance, respectively, while 0·8 is a value chosen to distinguish the very large positive and negative correlations. The variables are arranged so that strong positive correlations tend to lie near the principal diagonal and strong negative correlations tend to lie in the upper right and lower left corners.

The general similarity between the patterns for the SC and CC cases is not particularly surprising because many of the correlations are simply results of the definitions of the variables (a phenomenon previously encountered in the Horton analysis). For example, equation (11) gives relations between F, D, K, and $\bar{l}$ and equation (12) gives relations between F_e, F_i, D_e, D_i, K_e, K_i, λ, and α. Similarly, by neglecting the difference between n and $n - 1$, we have

$$\bar{l} = (\bar{l}_e + \bar{l}_i)/2$$

and

$$F = 2/(\bar{a}_e + \bar{a}_i) \tag{13}$$

while other more complicated relations can be derived as needed.

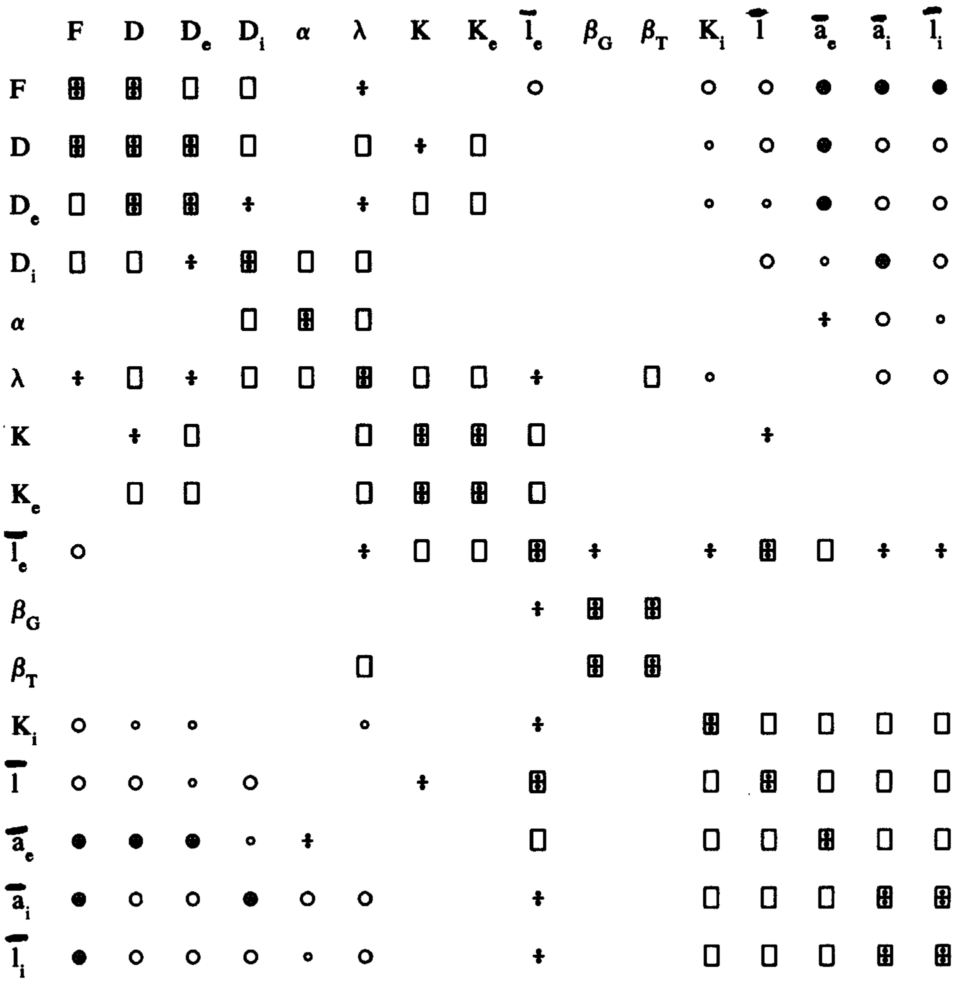
Although it is apparent that many of these pairs of variables should be strongly correlated, the exact nature of the pattern expected is certainly not obvious, and it is worth commenting on some of the details, particularly those which occur for both CC and SC networks. First of all, the patterns suggest a division of the variables into three groups such that the variables within each group tend to have rather strong positive correlations. In addition, groups I and III have rather strong negative correlations between their individual variables, while group II is weakly positively correlated with both of the others. The groups are:

Group I: F, D, D_e, D_i, α, λ

Group II: K, K_e, $\bar{l}_e$, β_G, β_T

Group III: K_i, $\bar{l}$, $\bar{a}_e$, $\bar{a}_i$, $\bar{l}_i$

In a number of instances, the quantitative agreement between the CC and SC results is remarkably good. For example, the highest positive correlation in both cases is that between $\bar{a}_i$ and $\bar{l}_i$ (CC, 0·954; SC, 0·968), while the exterior link lengths and link drainage areas have lower but comparable values of r_s (CC, 0·602; SC, 0·551). It is interesting to note that the presumably more objective method of determining exterior link lengths for the CC network did not result in any closer correlation between $\bar{l}_e$ and $\bar{a}_e$. The high correlation between $\bar{l}_i$ and $\bar{a}_i$ suggests that D_i, which is simply their ratio, should have a small coefficient of variation, and confirmation of this inference can be found in Tables XII and XIX (later in this section). D_i is however, strongly negatively correlated with both $\bar{l}_i$ and $\bar{a}_i$ and in fact the D_i–$\bar{a}_i$ coefficient is the most negative in both sets of data (CC, −0·896; SC, −0·908). In contrast with the D_i behaviour, D_e has a strong negative correlation with $\bar{a}_e$ but not with $\bar{l}_e$ (CC, −0·875 and −0·194; SC, −0·743 and 0·094).

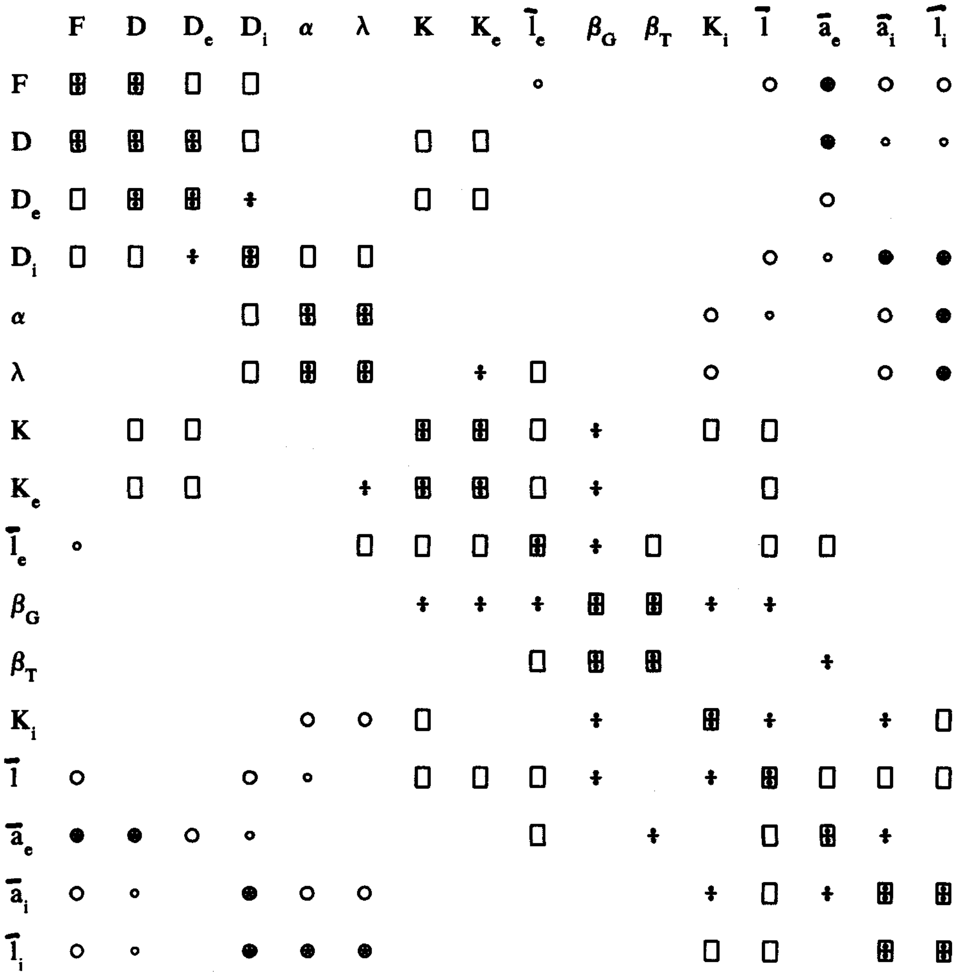


(a)

Figure 12. Pictorial representation of correlations between 16 watershed variables. Symbols indicate ranges of values for Spearman rank correlation coefficient as follows: domino, $1 \geqslant r_s > 0{\cdot}8$; rectangle, $0{\cdot}8 \geqslant r_s > 0{\cdot}456$; division sign, $0{\cdot}465 \geqslant r_s > 0{\cdot}306$; blank, $0{\cdot}306 \geqslant r_s > -0{\cdot}306$; small circle, $-0{\cdot}306 \geqslant r_s > -0{\cdot}465$; large circle, $-0{\cdot}465 \geqslant r_s > -0{\cdot}8$; filled circle, $-0{\cdot}8 \geqslant r_s \geqslant -1$

The shape factors β_G and $\hat{\beta}_T$ are different in character from the other 14 variables in that they cannot be directly expressed as functions of mean link lengths and mean link drainage areas. The main reason for including them here was to obtain a check on the sensitivity of the analysis. Their difference is indicated by the generally low correlation with the other variables; indeed, if only the CC data were considered, β_G and β_T would probably be identified as a separate group.

Despite the good agreement in correlation patterns, there are substantial differences in the compositions of the CC and SC networks. Some of these differences can be noted in Table XVI but others are not so obvious. For instance, examination of the table suggests that the K-factor is much the same for both methods since the mean of the ratio K_c/K_s is 1·15 and the coefficient of variation is 0·2. When the K-values are broken down into exterior and interior components, however, the apparent similarity disappears. Reference to Tables XII and XIX shows that in fact the roles of K_e and K_i are reversed in the two methods. Also the overall pattern of the interior–exterior ratios is quite different; for the CC case, we had $D_e/D_i \approx 1$, $K_e/K_i > 1$, and $F_e/F_i < 1$, while for the SC case all three ratios are about 0·66. Since $\lambda_s \approx 1$, equation (12) suggests that the ratios approximate the mean of $1/\alpha$, which is indeed 0·66.



(b)

Figure 12. (cont'd)

Table XIX. K, D, and F: exterior and interior components—CC method

Property	K_e	K_i	K_e/K_i	D_e (km^{-1})	D_i (km^{-1})	D_e/D_i	F_e (km^{-2})	F_i (km^{-2})	F_e/F_i
Mean	0·591	0·867	0·685	1·99	3·04	0·658	6·84	10.87	0·658
S.D.	0·127	0·081	0·150	0·30	0·31	0·098	1·42	2·49	0·191
C.V.	0·216	0·094	0·219	0·15	0·10	0·148	0·21	0·23	0·290
Max.	0·888	1·065	1·003	2·46	3·56	0·887	10·85	15·63	1·363
Min.	0·357	0·658	0·388	1·31	2·44	0·408	4·06	6·62	0·288

RÉSUMÉ AND COMMENTS

In the foregoing sections, a wide range of results have been presented in order to illustrate the scope and effectiveness of the new approaches to the analysis of drainage basin composition. Three of the discoveries stand out as especially important. First, the channel networks are slightly but significantly more elongated than predicted by the random model. This result is clearly established for the CC case and strongly suggested by the smaller set of data for the SC networks.

Second, the observed numbers of second-magnitude subnetworks are only 95 and 90 per cent of the expected values for the CC and SC cases, respectively. We can see qualitatively that this discrepancy is to some extent associated with the elongation properties. The disposition of the accompanying excess of *EI* links tends to produce a larger diameter than usual, an effect noted in the discussion of the topologic properties of low-magnitude CC networks. In order to obtain a quantitative assessment of this relationship, the factors f_2 (ratio of observed and expected numbers of second-magnitude subnetworks) and f_d (ratio of observed and expected values of the diameters) were computed. The Spearman rank correlation coefficients were $-0{\cdot}528$ and $-0{\cdot}384$ for the CC and SC networks, respectively. Both values are sufficiently large in magnitude to be significant at the 0·05 level but not nearly large enough to permit an acceptable quality of estimation. In our sample of 60, for example, there were several specific cases in which, contrary to the general correlation results, both f_2 and f_d were substantially greater than 1, or substantially less than 1.

Third, the length distribution for interior links depends upon the kind of link (interior or exterior) joined downstream. An analogous result was noted for exterior links, here confirming a previous discovery of Abrahams and Campbell (1976) although there are some differences of detail between our results and theirs. Also, as indicated in Tables X and XVIII, the distinctions hold for both SC and CC networks; in all four cases the *XI* mean link length is greater than the *XE* mean link length, the factors ranging from 10 to 50 per cent.

These three results not only provide important clues to the processes involved in the development of the Kentucky networks but they also raise some interesting subsidiary questions. For example, are the results peculiar to the eastern Kentucky region or are they also characteristic of other systems of natural networks? There are some scattered pieces of evidence which bear on the elongation property and the magnitude-2 abundances. Werner and Smart (1973, p. 286–287) reported the observed diameter distribution for 73 magnitude-10 dendritic channel networks from various parts of the eastern United States. Just as in the data of Table V, the random model can be accepted at the 0·05 level but the pattern of deviations shows too few small and too many large diameters. Smart (1968), in an early study of the topologic properties of natural channel networks, found that the observed abundance of second-order streams (or second-magnitude subnetworks) agreed well with the random model for streams in the eastern United States but was about 20 per cent too low for those in the western United States; the result thus suggests that the western networks are too elongated. Werner (private communication) substantiated this inference for the specific case of channel networks in the San Dimas National Forest, California; moreover, the departure from the random model was much more pronounced than for the comparable Kentucky data in Table VI.

In the most extensive study relating to this point, Abrahams (1977, in press) found a significant positive correlation between the abundance of Mock (1971) TS links and a relative relief parameter. In our terminology, a TS link is designated *EI* and the relative abundance is

$$I_1/N_1 = 1 - 2\,N_2/N_1 \tag{14}$$

Thus Abrahams' results indicate that the proportion of second-magnitude subnetworks decreases with increasing relative relief, and he suggests that such an effect can account for the difference noted by Smart between streams in the eastern and western United States. Taking all of these results into account, it seems plausible to speculate that network elongation generally increases with increasing relative relief (or slope) and that the Kentucky data marks a specific point in the relationship. Because of the rather tenuous connection between elongation and abundance of second-magnitude links, however, more data taken under other conditions is required to prove or disprove the speculation.

The only other available data relating to *XI* and *XE* link lengths are the previously mentioned results of Abrahams and Campbell (1976). They measured *EI* and *EE* link lengths in six drainage systems in Australia and found significant differences in the two distributions in all cases. The *EI* links tended to be longer, with the ratio $\bar{l}_{II}/\bar{l}_{IE}$ falling in the range 1·12–1·63. They attributed this difference to an increase in *EI* link length as the valleys became larger downstream and showed that *EI* links that joined links of magnitude $\mu < 5$ had length distributions much like the *EE* links and were appreciably shorter than the *EI* links which joined links of magnitude $\mu \geqslant 5$. This latter phenomenon does not occur in our results; the difference in mean length for the two groups is less than one per cent. This result is consistent with the fact that the Kentucky maps do not indicate much tendency for the sub-basins which control the length of *EI* links to widen downstream. As demonstrated by the example of Blacklog Fork in Figure 11a, to a rough approximation the drainage basins are formed by combining rectangular sub-basins whose magnitudes lie in the range 5–20.

In order to provide some more data for comparison I have made link length measurements on maps of channel networks published by Melton (1957), Coates (1958), and Maxwell (1960). The Melton and Maxwell maps were prepared from field studies and the Coates maps from a combination of field, topographic map, and aerial photograph studies. I have also made measurements on channel networks outlined by the CC method on U.S.G.S. 1:24,000 topographic maps of southern Indiana (Nashville quadrangle), the Missouri Ozarks (Waynesville quadrangle), and the Great Smoky Mountains (Thunderhead, N. Car.–Tenn., quadrangle). The ratios of *XI* to *XE* mean link lengths all fell in the range 1·15–1·81. As the sample sizes were typically about 100, these ratios are not particularly well established but for all except the lowest value they were significantly (0·05 level) greater than 1. Although clearly much more detailed study is needed, these preliminary results taken from a rather wide range of environments certainly suggests that the link length distribution properties observed for the Kentucky region is a characteristic feature of natural dendritic networks.

One puzzling aspect of the Kentucky data is the qualitative, and to some extent quantitative, similarity in the properties of the CC and SC networks. Not only are they alike in the three important results discussed above but also in some other features, such as the cis–trans ratio and the correlations illustrated in Figure 12. It seems doubtful that this level of agreement could be accidental. On the other hand, there is nothing special about the SC slope criterion for sources; are we then to assume that any other reasonably objective and consistent rule for identifying sources would produce still another set of channel networks with the same general properties? It is perhaps possible that as a landscape evolves in drainage density and relief, the same set of influences act at all scales, thus producing at any instant channel networks that have the general kinds of diameter and link length distributions observed for the Kentucky area. What I have done here, of course, is to reintroduce, with a somewhat different interpretation, Horton's (1945) suggestion that channel networks outlined on small scale maps have a number of features of composition in common with the more detailed networks outlined on large-scale maps of the same basins. Regardless of the validity of this speculation, one empirical consequence of the general similarity of the CC and SC networks and of the similarity of the link length distributions obtained by the CC method and by field studies is the indication that the often-criticized CC results are in good contact with geomorphic reality.

As mentioned in earlier sections, a number of the topologic and geometric properties of the Kentucky networks deviate considerably from the two postulates of the random model. Perhaps the most important discrepancy occurs in the link length distributions, both in the internal division into *XE* and *XI* groups and in the rather large variation in link lengths from one network to another in the region. Smart and Werner (1976) have, however, previously used this same data to demonstrate the effectiveness of the random model in estimating, explaining, and predicting geomorphic properties. Although it may appear on first sight that the two sets of results are incompatible, Smart and Werner (1967, p. 231–232) have already noted that such applications of the random model are, for reasons not completely understood, generally rather insensitive to the kinds of deviations which have shown up in this detailed study. As the Smart and Werner results have their own empirical justification, there seems to be no reason to consider the two papers in conflict. The results of this analysis, however, do emphasize the necessity for caution and checking in extending and generalizing the Smart–Werner applications for this or any other set of network data.

Although I feel that the usefulness of these new methods of analysis in fluvial geomorphology is now well established, their role in hydrology is still in question. That there should be some relation between network composition and hydrologic properties seems intuitively apparent. Many methods of synthesizing hydrographs require at least some information about topologic and geometric properties and some methods, such as those of Surkan (1969, 1974) and Delleur and Lee (1973), require a detailed and explicit description. More recently, Kirkby (1976) has given a comprehensive theoretical treatment of the problem. Attempts to provide a quantitative demonstration of such effects by observations have, however, not fared very well. Several authors have looked for an effect on hydrologic properties, particularly hydrograph parameters, of some of the variables listed in Table II, but the results have been little, if any, more promising than the corresponding studies for the Horton ratios. Benson (1962), in a study of discharge records of 164 gauging stations in humid regions of the United States, found a significant correlation between peak discharge and both the geometric shape factor and 'index of drainage pattern' which measured the degree of elongation of the network. Partial correlation studies showed, however, that these effects disappeared when the associations of all variables with area and mainstream slope were taken into account. Similar results (Benson (1964)) were obtained for floods in the southwestern United States. Rice (1970) considered mainstream length and flowpath coefficient of variation as part of a group of 24 watershed variables in a study of hydrographs of chaparral watersheds in the San Dimas National Forest, California; both variables were however deleted for lack of significant independent effect before the final stages of the analysis were reached. Rogers (1972) attempted to predict hydrograph shape from the distribution of flowpath lengths. While agreeing that there is undoubtedly some relation between the two quantities, I do not, for reasons explained below, believe that it is as close as Rogers claims, and I did not find his examples particularly convincing.

The major difficulty in obtaining unambiguous quantitative evidence of the effects of network composition on hydrologic properties is that of disentangling them from the influences of a number of other complex phenomena. To take a specific example, a partial list of the factors which may be expected to have a significant influence on the shape of a storm hydrograph are: (1) rainstorm intensity as a function of both time and location, (2) soil moisture and groundwater conditions at the beginning of the rainfall, (3) partial-areas effects (which may to some extent depend on the first two factors), and (4) network composition, including channel slopes. The labour and instrumentation required to adequately characterize the first two factors is in most cases prohibitive; instead they are usually represented by a few intuitively-selected empirical parameters based on a limited number of measurements. Even with this simplification, the separation of the various effects remains a formidable problem in statistical analysis and one which has not, to my knowledge, been carried out with any degree of success. These observations are of course in no way original; detailed discussions of the points raised here can be found in many places in the literature, e.g., Anderson (1949), Rice (1970). They are repeated here simply to re-emphasize the complex nature of the problem and the rather gloomy prospects for its solution.

In closing the comments, it seems worthwhile to point out that the multiplicity and range of results obtained in these investigations were possible because of the data-sorting ability inherent in the use of link data vectors. In reviewing the sections on data analysis, we find that subnetworks were sorted according to (1) magnitude, (2) diameter, (3) ambilateral class, and (4) type of link joined downstream. Interior link lengths were sorted according to (1) link magnitude, (2) size with respect to the median, (3) type of link joined downstream, and (4) cis–trans classification. Exterior link lengths were sorted according to type of link joined downstream and the *EI* subgroup according to the magnitude of link joined. There were also a number of examples of double sorting procedures to construct 2×2 contingency tables. All but one of the sorting criteria listed above are topologic. The algorithms required for this work were quite simple and in most cases were based on the use of the magnitude vector to define subnetworks and to locate the right forks of interior links; an example was given in the discussion of the path length algorithm.

Finally, I would like to acknowledge the excellent assistance of Thomas McElroy in collecting and coding the data and in carrying out a considerable portion of the analytical work. I am also indebted to C. Werner for helpful comments on an early draft of the manuscript and to A. D. Abrahams for a preprint of his paper on the relation between relative relief and the abundance of TS links.

REFERENCES

Abrahams, A. D. (1972). 'Factor analysis of drainage basin properties: evidence for stream abstraction accompanying degradation of relief', *Water Resources Research*, **8**, 624–633.

Abrahams, A. D. (1977). 'The factor of relief in the evaluation of channel networks in mature drainage basins', *American Journal of Science* (in press).
Abrahams, A. D. and Campbell, R. N. (1976). 'Source and tributary-source link lengths in natural channel networks', *Bulletin of the Geological Society of America*, **87**, 1016–1020.
Anderson, H. W. (1949) 'Flood frequencies and sedimentation from forest watersheds', *Transactions of the American Geophysical Union*, **30**, 567–584.
Anderson, H. W. (1950). 'Discussion of "Flood frequencies and sedimentation from forest watersheds" ', *Transactions of the American Geophysical Union*, **31**, 621–623.
Anderson, H. W. and Trobitz, H. K. (1949). 'Influence of some watershed variables on a major flood', *Journal of Forestry*, **47**, 347–356.
Benson, M. A. (1962). 'Factors influencing the occurrence of floods in a humid region of diverse terrain', *U.S. Geological Survey Water-Supply Paper 1580-B*.
Benson, M. A. (1964). 'Factors affecting the occurrence of floods in the southwest United States', *U.S. Geological Survey Water-Supply Paper 1580-D*.
Bowden, K. L. and Wallis, J. R. (1964). 'Effect of stream-ordering technique on Horton's laws of drainage composition', *Bulletin of the Geological Society of America*, **75**, 767–74.
Broscoe, A. J. (1959). 'Quantitative analysis of longitudinal profiles of small watersheds', *Project NR 389-042 Technical Report No. 18*, Department of Geology, Columbia University.
Chorley, R. J. (1957). 'Illustrating the laws of morphometry', *Geological Magazine*, **94**, 140–150.
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.
Cochran, W. G. (1952). 'Some methods for strengthening the common χ^2 tests', *Biometrics*, **10**, 417-451.
Dacey, M. F. and Krumbein, W. C. (1976). 'Three growth models for stream channel networks', *Journal of Geology*, **84**, 153–163.
Delleur, J. W. and Lee, M. T. (1973). 'A rainfall-runoff model based on the watershed stream network', *Proccedings of the Symposium on the Design of Water Resources Projects with Inadequate Data, Vol. I*, Madrid, 75–87.
Eyles, R. J. (1968). 'Stream net ratios in west Malaysia', *Bulletin of the Geological Society of America*, **79**, 701–712.
Flint, J. J. (1976). 'Link slope distribution in channel networks', *Water Resources Research*, **12**, 645–654.
Hoel, P. G. (1962). *Introduction to Mathematical Statistics*, 3rd edition, New York, Wiley.
Horton, R. E. (1945). 'Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology', *Bulletin of the Geological Society of America*, **56**, 275–370.
James, W. R. and Krumbein, W. C. (1969). 'Frequency distributions of stream link lengths', *Journal of Geology*, **77**, 544–565.
Jarvis, R. S. (1972). 'A 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*, University of Cambridge.
Jarvis, R. S. (1976a). 'Stream orientation structures in drainage networks', *Journal of Geology*, **84**, 563–582.
Jarvis, R. S. (1976b). 'Classification of nested tributary basins in analysis of drainage basin shape', *Water Resources Research*, **12**, 1151–1164.
Jarvis, R. S. (1976c). 'Link length organization and network scale dependencies in the network diameter model', *Water Resources Research*, **12**, 1215–1225.
Jarvis, R. S. and Werritty, A. (1975). 'Some comments on testing random topology stream network models', *Water Resources Research*, **11**, 309–318.
Kirkby, M. J. (1976). 'Tests of the random network model and its application to basin hydrology', *Earth Surface Processes*, **1**, 197–212.
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.
Leopold, L. B. and Langbein, W. B. (1962) 'The concept of entropy in landscape evolution', *U.S. Geological Survey Professional Paper 500-A*.
Leopold, L. B. and Miller, J. P. (1956). 'Ephemeral streams-hydraulic factors and their relation to the drainage net', *U.S. Geological Survey Professional Paper 282–A*.
Maxwell, J. C. (1960). 'Quantitative geomorphology of the San Dimas Experimental Forest, California', *Project NR 389–042 Technical Report No. 19*, Department of Geology, Columbia University.
McFarlan, A. C. (1951). *Geology of Kentucky*, University of Kentucky, Lexington, Kentucky.
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.
Milling, M. E. and Tuttle, S. D. (1964). 'Morphometric, study of two drainage basins near Iowa City, Iowa', *Proceedings of the Iowa Academy of Sciences*, **71**, 304–319.
Mock, S. J. (1971). 'A classification of channel links in stream networks', *Water Resources Research*, **7**, 1558–1566.
Morisawa, M. E. (1962). 'Quantitative geomorphology of some watersheds in the Appalachian Plateau', *Bulletin of the Geological Society of America*, **73**, 1025–1046.
Price, W. E., Jr., Mull, D. S. and Kilburn, C. (1962). 'Reconnaissance of groundwater resources in the eastern coal field region, Kentucky', *U.S. Geological Survey Water Supply Paper No. 1607*.
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. and Langbein, W. B. (1966). 'Probability concepts in geomorphology', *U.S. Geological Survey Professional Paper No. 500-C*.
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.
Siegel, S. (1956). *Nonparametric Statistics for the Behaviourial Sciences*, New York, McGraw-Hill.
Smart, J. S. (1968). 'Statistical properties of stream lengths', *Water Resources Research*, **4**, 1001–1014.
Smart, J. S. (1969a). 'Topological properties of channel networks', *Bulletin of the Geological Society of America*, **80**, 1757–1774.
Smart, J. S. (1969b). 'Distribution of interior link lengths in natural channel networks', *Water Resources Research*, **5**, 1337–1342.
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 Hydroscience*, **8**, ed. V. T. Chow, New York, Academic Press, p. 305–346.
Smart, J. S. (1973). 'The random model in fluvial geomorphology', *Proceedings of the Fourth Annual Geomorphology Symposium*, ed. Marie Morisawa, Binghamton, New York, p. 27–49.
Smart, J. S. (1976). 'Joint distribution functions for link lengths and drainage areas', *Random Processes in Geology*, ed. D. F. Merriam, New York, Springer-Verlag, p. 112–123.
Smart, J. S., and Werner, C. (1976). 'Applications of the random model of drainage basin composition', *Earth Surface Processes*, **1**, 219–233.
Strahler, A. N. (1952). 'Hypsometric (area-altitude) analysis of erosional topography', *Bulletin of the Geological Society of America*, **63**, 1117–1142.
Strahler, A. N. (1957). 'Quantitative analysis of watershed geomorphology', *Transactions of the American Geophysical Union*, **38**, 913–920.
Strahler, A. N. (1958). 'Dimensional analysis applied to fluvially eroded landforms', *Bulletin of the Geological Society of America*, **69**, 279–300.
Surkan, A. J. (1969). 'Synthetic hydrographs: effects of network geometry', *Water Resources Research*, **5**, 112–128.
Surkan, A. J. (1974). 'Simulation of storm velocity effects on flow from distributed channel networks', *Water Resources Research*, **10**, 1149–1160.
Surkan, A. J. and Kelton, P. (1974). 'The simulation of urban hydrologic transfer functions using a binary tree model', *International Journal of Systems Science*, **5**, 639–654.
Werner, C. and Smart, J. S. (1973). 'Some new methods of topologic classification of channel networks', *Geographical Analysis*, **5**, 271–295.
Woodruff, J. F. (1964). 'A comparative analysis of selected drainage basins', *The Professional Geographer*, **16(4)**, 15–19.
Yang, C. T. and Stall, J. B. (1971). 'A note on the map scale effect in the study of stream morphology', *Water Resources Research*, **7**, 709–712.