Statistical Properties of Stream Lengths

J. S. SMART

IBM Watson Research Center, Yorktown Heights, New York 10598

Abstract. Two basic assumptions are employed in this treatment of the statistics of stream lengths: (1) All topologically distinct networks with a given number of sources are equally likely (after Shreve). (2) Lengths of interior links in a given network are independent random variables drawn from the same population. The mathematical development leads to an approximate expression for $\bar{L}_\omega$ that contains no adjustable parameters and that depends only on the stream numbers and the mean link length. This expression gives somewhat better agreement with data on actual stream systems than does Horton's law of stream lengths; other advantages over Horton's law are cited. Quantitatively, our procedure appears to account for about 65% of the variance in mean stream length data for third- and fourth-order basins. If exact values of the link numbers are introduced into the calculation, the unexplained variance is reduced to about 15%. For a complete statistical description of stream lengths, it is necessary to know the distribution of interior link lengths. Results of computer simulation studies suggest that this distribution is negative exponential. Data taken on two small watersheds (140 links) show reasonable agreement with this hypothesis. (Key words: Geomorphology; rivers; drainage basin characteristics)

INTRODUCTION

A famous paper by *Horton* [1945] laid the oundation for much of the subsequent work n quantitative geomorphology of drainage asins. In particular, Horton made two major ontributions to the study of stream patterns. irst, he devised a system of stream classifica-on, or ordering, which proved to be very eful in the quantitative discussion of drain-e basin composition. Horton's ordering sys-m and a modification proposed later by *ahler* [1952] are both well known and need description here. *Shreve* [1966] gives an ellent discussion of the relation between the systems. In this paper, we shall follow example of most recent workers and use Strahler system exclusively. Terms such 'stream,' 'stream number,' and 'stream gth' are to be interpreted according to ahler ordering unless otherwise specified. Iorton's second contribution was his two s of drainage composition:

w of Stream Numbers

$$N_\omega \approx R_B^{(\Omega-\omega)} \qquad 1 \leq \omega \leq \Omega \qquad (1)$$

ere N_ω is the number of streams of order n a basin of order Ω, and R_B is the bifurca-n ratio; and

Law of Stream Lengths

$$\bar{L}_\omega \approx R_L^{(\omega-1)} \bar{L}_1 \qquad (2)$$

where $\bar{L}_\omega$ is the mean length of streams of order ω, and R_L is the stream length ratio.

Horton also implied that there should be an analogous relation for basin areas, and such a law was later stated explicitly by *Schumm* [1956, p. 606]:

Law of Basin Areas

$$\bar{A}_\omega \approx R_A^{(\omega-1)} \bar{A}_1 \qquad (3)$$

where $\bar{A}_\omega$ is the mean area drained by streams of order ω, and R_A is the basin area ratio.

The laws are statistical in nature; that is, they are not intended to provide exact descriptions of individual drainage basins but rather to give the average behavior, or central tendency, for a large number of basins. The values of the parameters R_B, R_L, and R_A assigned to a particular basin are generally determined by plotting N_ω, $\bar{L}_\omega$, and $\bar{A}_\omega$ versus ω on semilog paper and determining the 'best fit' straight lines by least-squares analysis; the slopes of the lines are then $-\log R_B$, $\log R_L$, and $\log R_A$ for equations 1, 2, and 3, respectively. Values of R_B in actual stream systems typically range between 2.5 and 5.0, values of R_L between 1.5 and 3.0, and values of R_A between 3.5 and 6.0.

Statistical Properties of Stream Lengths

J. S. SMART

IBM Watson Research Center, Yorktown Heights, New York 10598

Abstract. Two basic assumptions are employed in this treatment of the statistics of stream lengths: (1) All topologically distinct networks with a given number of sources are equally likely (after Shreve). (2) Lengths of interior links in a given network are independent random variables drawn from the same population. The mathematical development leads to an approximate expression for $\bar{L}_\omega$ that contains no adjustable parameters and that depends only on the stream numbers and the mean link length. This expression gives somewhat better agreement with data on actual stream systems than does Horton's law of stream lengths; other advantages over Horton's law are cited. Quantitatively, our procedure appears to account for about 65% of the variance in mean stream length data for third- and fourth-order basins. If exact values of the link numbers are introduced into the calculation, the unexplained variance is reduced to about 15%. For a complete statistical description of stream lengths, it is necessary to know the distribution of interior link lengths. Results of computer simulation studies suggest that this distribution is negative exponential. Data taken on two small watersheds (140 links) show reasonable agreement with this hypothesis. (Key words: Geomorphology; rivers; drainage basin characteristics)

INTRODUCTION

A famous paper by *Horton* [1945] laid the oundation for much of the subsequent work n quantitative geomorphology of drainage asins. In particular, Horton made two major ontributions to the study of stream patterns. irst, he devised a system of stream classificaon, or ordering, which proved to be very eful in the quantitative discussion of draine basin composition. Horton's ordering sysm and a modification proposed later by *rahler* [1952] are both well known and need description here. *Shreve* [1966] gives an cellent discussion of the relation between the systems. In this paper, we shall follow example of most recent workers and use Strahler system exclusively. Terms such 'stream,' 'stream number,' and 'stream gth' are to be interpreted according to ahler ordering unless otherwise specified. Horton's second contribution was his two s of drainage composition:

w of Stream Numbers

$$N_\omega \approx R_B^{(\Omega-\omega)} \qquad 1 \leq \omega \leq \Omega \qquad (1)$$

ere N_ω is the number of streams of order n a basin of order Ω, and R_B is the bifurcan ratio; and

Law of Stream Lengths

$$\bar{L}_\omega \approx R_L^{(\omega-1)} \bar{L}_1 \qquad (2)$$

where $\bar{L}_\omega$ is the mean length of streams of order ω, and R_L is the stream length ratio.

Horton also implied that there should be an analogous relation for basin areas, and such a law was later stated explicitly by *Schumm* [1956, p. 606]:

Law of Basin Areas

$$\bar{A}_\omega \approx R_A^{(\omega-1)} \bar{A}_1 \qquad (3)$$

where $\bar{A}_\omega$ is the mean area drained by streams of order ω, and R_A is the basin area ratio.

The laws are statistical in nature; that is, they are not intended to provide exact descriptions of individual drainage basins but rather to give the average behavior, or central tendency, for a large number of basins. The values of the parameters R_B, R_L, and R_A assigned to a particular basin are generally determined by plotting N_ω, $\bar{L}_\omega$, and $\bar{A}_\omega$ versus ω on semilog paper and determining the 'best fit' straight lines by least-squares analysis; the slopes of the lines are then $-\log R_B$, $\log R_L$, and $\log R_A$ for equations 1, 2, and 3, respectively. Values of R_B in actual stream systems typically range between 2.5 and 5.0, values of R_L between 1.5 and 3.0, and values of R_A between 3.5 and 6.0.

The Horton analysis may be regarded as an attempt to give a quantitative description of drainage basin composition in terms of five parameters: R_B, R_L, R_A, $\bar{L}_1$, and $\bar{A}_1$, or some equivalent set [*Horton*, 1945, p. 295]. Once these five numbers are known, approximate values for other geomorphic parameters, such as total stream length, total basin area, drainage density, and stream frequency, can easily be obtained. *Leopold and Miller* [1956] have shown how the analysis can be extended to predict width, depth, discharge, sediment load, and other hydraulic variables. Although the Horton and Strahler ordering techniques have been much used by hydrologists and geomorphologists, they have also attracted a certain amount of criticism. In my opinion, the three most relevant objections are those mentioned by *Scheidegger* [1965]: (1) both basin order and the order of individual streams depend on the scale of map used; (2) the order in a stream network changes only when two streams of equal order join, whereas the hydrologic properties of a stream change at any kind of junction; (3) the rule for combining stream numbers is not distributive; symbolically, $2 + (1 + 1) = 3$ but $(2 + 1) + 1 = 2$. It is true that a certain amount of the scatter observed in geomorphic data on drainage basins is directly traceable to these ambiguities and inconsistencies in the ordering scheme. However, it appears somewhat unreasonable to draw from this fact the conclusion that stream ordering is an unrewarding exercise. Despite its obvious deficiencies, the Horton analysis, with just a few numbers, still suceeds in providing a surprising amount of information about a very complicated system.

Several authors, e.g., *Morisawa* [1962], have noted that by eliminating ω from equations 1–3 one can obtain direct power function relations between any two of the quantities N_ω and $\bar{L}_\omega$ and $\bar{A}_\omega$, the exponents being ratios of the logarithms of the branching ratios. Data on drainage basin parameters also exhibit another, more subtle relation, which has apparently not been mentioned in the literature: the branching ratios R_B, R_L, and R_A are rather highly correlated. As examples, Figure 1 shows R_L versus R_B ($r = 0.726$) for 46 fourth-order basins studied by *Melton* [1957], and Figure 2 shows R_A versus R_B ($r = 0.899$) for 12 Appalachian region watersheds investigated by

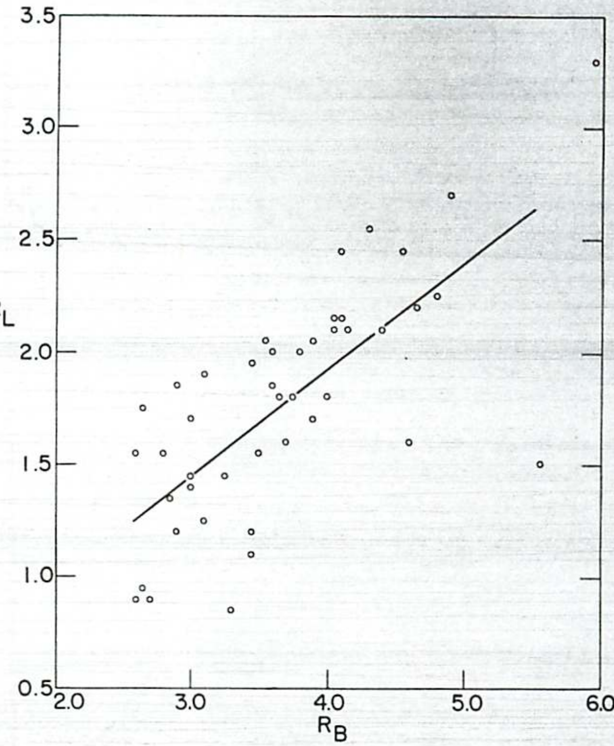

Fig. 1. Regression of R_L on R_B for 46 fourth-order drainage basins. Data from *Melton* [1957].

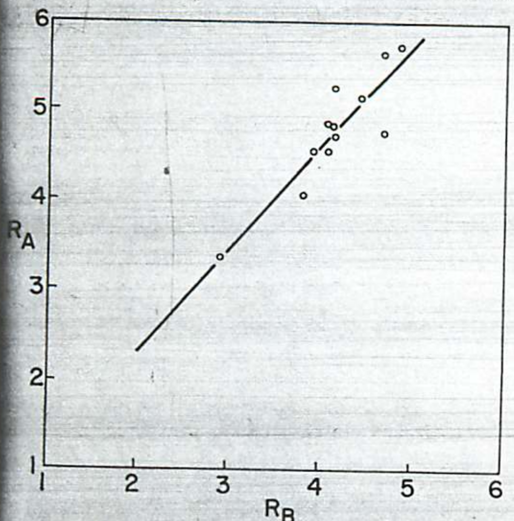

Fig. 2. Regression of R_A on R_B for twelve Ap-
palachian region drainage basins. Data from
Morisawa [1962].

lorisawa [1962]. The correlation illustrated in igure 2 is especially striking, because it includes ata on basins of different order (4–7) and on ncomplete basins for which $A_\Omega/\bar{A}_{\Omega-1}$ is unusually mall. Examination of other published data on ranching ratios shows positive correlation coeffi- ients ranging between 0.4 and 0.9.

This result indicates that the geomorphic arameters of a drainage basin are not inde- endent of each other but rather are related in subtle fashion. From a practical point of ew, it suggests the possibility of giving a antitative descriptive of drainage basin com- sition with fewer parameters than are re- ired by a Horton analysis. From a funda- ntal point of view, it encourages theoretical estigations to determine the exact nature the interrelations of various parameters.

ne important step in this direction has eady been taken by *Shreve* [1966]. He pro- ed a theory of network patterns based on assumption that all topologically distinct works with a given number of sources are ally likely. With this model, he was able explain some of the observed properties of ual networks, such as the fact that $N_\omega/N_{\omega+1}$ about 4 and the fact that Horton plots of eam numbers tend to be concave upwards. eve also showed that stream numbers for ual drainage basins had a distribution that agreed reasonably well with the theoretical one and suggested that his model could serve as a starting point for derivations of the metrical properties of drainage basins, such as lengths and areas.

Any approach to the problem of stream lengths should logically start with a consideration of the properties of links. A *link,* as defined by *Shreve* [1966, p. 20; 1967, pp. 178–179], is a section of channel network between two successive junctions, between a source and the first junction downstream, or between the outlet and the first junction upstream. Links that have one end point at a source are called exterior links and are identical with first-order Strahler streams; all other links are called interior links, and each interior link is part of a higher order Strahler stream. Networks with N_1 sources have N_1 exterior links and N_1–1 interior links. *Shreve* [1967] has proposed a method of classifying links which is analogous to the Horton and Strahler schemes for stream ordering. All exterior links have magnitude 1; the magnitude of an interior link is the sum of the magnitudes of the two links that contribute to it.

In some ways, link classification gives a simpler and more precise description of drainage basin composition than does stream ordering. For example, a basin with N_1 sources always has magnitude N_1, but its order is one of a number of possible values. Similarly, a network with N_1 sources always has $2N_1 - 1$ links, but the number of streams can take on any integral value between $(N_1 + 2^{\Omega-1} - 1)$ and $((2\Omega - 1)\ \Omega^{-1}N_1 + 1)$. Moreover, although the point has not been checked, it appears that the link magnitude number should give better quality hydrologic information than does the stream order number. A link of magnitude μ carries the discharge from μ sources plus the overland flow from areas contiguous to $\mu 2 - 1$ links, whereas the corresponding relations for stream orders are again indefinite.

There are, however, some disadvantages in using links. Figure 3 gives a comparison of stream ordering and link classification for the channel network of a basin with 31 sources (headwaters of Gourd Creek, Yancy Mills quadrangle, Missouri). Values of N_ω, the number of streams of order ω, and n_μ, the number of links of magnitude μ, are given in Table 1.

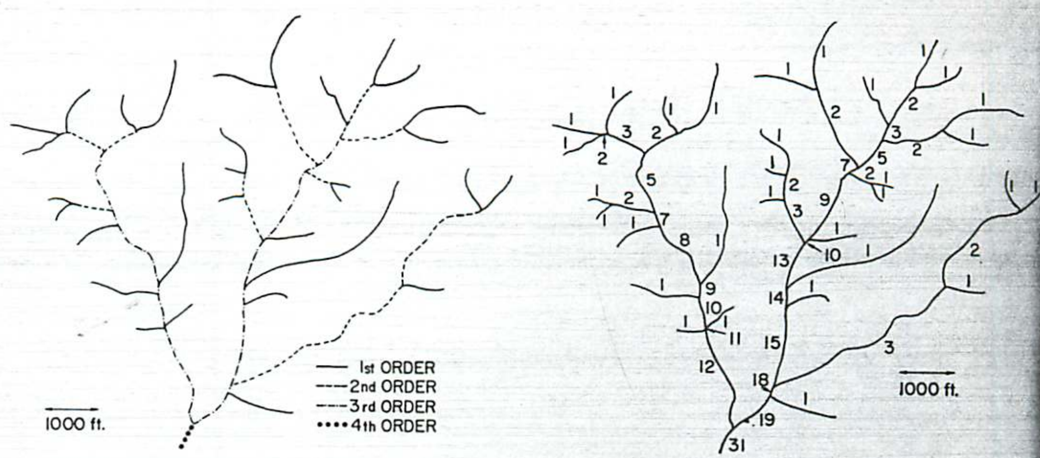

Fig. 3. Left, Channel network for headwaters of Gourd Creek (Mo.) with Strahler ordering. Right, Same network with Shreve classification of links.

With link classification, 31 numbers are required to characterize the networks, whereas only 4 are needed with stream ordering; this ratio will of course be maintained when length and area data are added. Even though 15 of the 31 link numbers are zero, the precision gained with the link classification is paid for by increased bookkeeping, and the cost will rise sharply as the number of sources increases.

Regardless of this and other possible difficulties in practical usage, from a theoretical viewpoint the link is the basic unit from which channel networks are constructed. Consequently, any discussion of lengths of more complex channel segments, such as Strahler streams, should be based on a model of link length distribution. In the next section, we show how combining Shreve's theory with a very general assumption about the nature of the link length distribution leads to a simple approximate relation between stream numbers and mean stream lengths.

TABLE 1. Classification Systems for Gourd Creek (Mo.)

Strahler streams								
	ω	1	2	3	4			
	N_ω	31	9	2	1			
Shreve links								
μ	1	2	3	4	5	6	7	8
n_μ	31	9	4	0	2	0	2	1
μ	9	10	11	12	13	14	15	16
n_μ	2	2	1	1	1	1	1	0
μ	17	18	19	20	21	22	23	24
n_μ	0	1	1	0	0	0	0	0
μ	25	26	27	28	29	30	31	
n_μ	0	0	0	0	0	0	1	

STRAHLER STREAMS

In considering the question of length distributions for Strahler streams of order $\omega \geq 2$, two related but distinct problems can be identified. One is how the total length of such streams is divided among the various orders, and the other is how the total length for a given order is divided among the individual streams of that order. A solution of the first problem will provide information about $\bar{L}_\omega$, the mean length of streams of order ω in a particular network. As practically all of the length data for actual stream systems consists of values of $\bar{L}_\omega$, our initial analysis is confined to this question. The first step, therefore, is to obtain an explicit expression for $P(\mathfrak{L}; \omega, \mathfrak{N}_\Omega) d\mathfrak{L}$, the probability that streams of order ω in a network with stream numbers $\mathfrak{N}_\Omega = N_1, N_2, \cdots, N_{\Omega-1}, 1$ have a total length between $\mathfrak{L}$ and $\mathfrak{L} + d\mathfrak{L}$.

Strahler streams of order $\omega \geq 2$ are composed of sequences of interior links. Consequently, the length distribution function for a particular order will depend not only on the distribution function for individual link lengths but also on the way in which the $N_1 - 1$ interior links are divided among the various orders. This consideration suggests that $P(\mathfrak{L}; \omega, \mathfrak{N}_\Omega)$ can be expressed as a compound distribution, namely, as a product of the probability that streams of order ω and ν links and the probability that ν links have length $\mathfrak{L}$. The quantitative formulation of this idea is

$$P(\mathcal{L}; \omega, \mathfrak{N}_\Omega) = \sum_\nu f(\nu; \omega, \mathfrak{N}_\Omega) g(\mathcal{L}; \nu, \omega, \mathfrak{N}_\Omega) \quad (4)$$

where $f(\nu; \omega, \mathfrak{N}_\Omega)$ is the probability that streams of order ω in a network with stream numbers $\mathfrak{N}_\Omega$ ave a total of exactly ν links, and $g(\mathcal{L}; \nu, \omega, \mathfrak{N}_\Omega)$ s the probability density function for the length $\mathcal{L}$ of a set of ν links of order ω in a network with tream numbers $\mathfrak{N}_\Omega$. The summation is carried ut over all possible values of ν; in general, the mits of summation will depend on $\mathfrak{N}_\Omega$. Both $(\nu; \omega, \mathfrak{N}_\Omega)$ and $g(\mathcal{L}; \nu, \omega, \mathfrak{N}_\Omega)$ may depend on arameters other than those listed in the formula.

It should also be noted that for any particular etwork the numbers ν_ω of links for different rders are not independent but must satisfy the ondition

$$\sum_{\omega=2}^{\Omega} \nu_\omega = N_1 - 1 \quad (5)$$

Up to this point, the development has been ompletely general. We now introduce two basic ssumptions that determine the properties of $(\nu; \omega, \mathfrak{N}_\Omega)$ and $g(\mathcal{L}; \nu, \omega, \mathfrak{N}_\Omega)$ and enable us to btain explicit results:

1. All topologically distinct networks with a iven number of sources are equally likely *Shreve*, 1967].

2. Lengths of interior links in a given network re independent random variables drawn from he same population.

Note that these assumptions are independent the ordering scheme and could be used to disss stream length statistics in any system of dering.

Although the second assumption may not ppear very restrictive, it does have some imporant effects on the nature of $g(\mathcal{L}; \nu, \omega, \mathfrak{N}_\Omega)$. First all, it means that g is independent of ω, i.e., ks of different orders have the same length stribution. More important, as the individual k lengths are independent random variables, e mean length of ν links is just ν times the mean ngth of individual links. In this case the expreson for the mean value of $\mathcal{L}$ becomes particularly mple.

$$\begin{aligned} E(\mathcal{L}; \omega, \mathfrak{N}_\Omega) &= \int_0^\infty \mathcal{L} P(\mathcal{L}; \omega, \mathfrak{N}_\Omega)\, d\mathcal{L} \\ &= \bar{l}_i \sum_\nu \nu f(\nu; \omega, \mathfrak{N}_\Omega) \\ &= \bar{\nu} \bar{l}_i \end{aligned} \quad (6)$$

where $\bar{l}_i$ is the mean length of links in the population. The variance and other moments will in general depend on the explicit form of $g(\mathcal{L}; \nu, \mathfrak{N}_\Omega)$.

Whereas the second assumption requires only that $g(\mathcal{L}; \nu, \mathfrak{N}_\Omega)$ be one of a rather wide class of functions, the function $f(\nu; \omega, \mathfrak{N}_\Omega)$ is completely determined by the first assumption. In a qualitative way, we can see that $f(\nu; \omega, \mathfrak{N}_\Omega)$ is closely related to the topologic properties of a channel network by the fact that the number of links in a stream is determined by the number of tributaries it has. This observation suggests that, for a given network, stream numbers and stream lengths should be related; it explains, at least qualitatively, the observed correlation of R_B and R_L (Figure 1).

As a specific example of how $f(\nu; \omega, \mathfrak{N}_\Omega)$ is related to the channel pattern, consider a fourth-order network with $N_1 = 20$, $N_2 = 8$, $N_3 = 3$, $N_4 = 1$. Two of the three third-order streams are required to create the fourth-order stream, and the remaining one must join it at some interior point, as shown in Figure 4. Thus the fourth-order stream will have at least two links.

Six of the eight second-order streams must join in pairs to form the three third-order streams; the other two can be distributed among the five links of third and fourth order in 15 possible ways. Finally, the network is completed by using 16 first-order streams to form the eight second-order

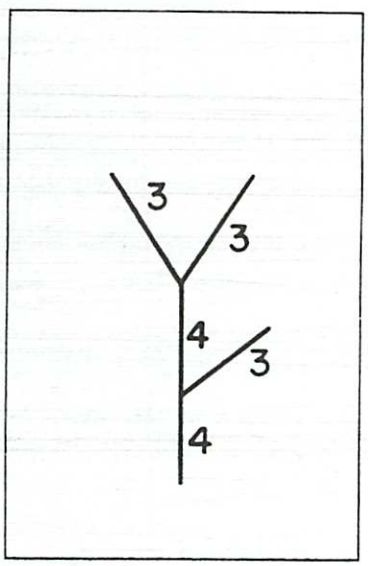

Fig. 4. Arrangement of third and fourth order streams.

streams and by distributing the remaining 4 first-order streams among the 15 interior links of higher order; this last step can be performed in 3060 distinguishable ways.

As each order is added to the network, we must assign $R_\omega = N_\omega - 2N_{\omega+1}$ tributaries or 'excess' streams of order ω to $2N_{\omega+1} - 1$ links of higher order. This is just the problem of putting r identical objects into n cells; the number of different ways in which it can be done is known from elementary combinatorial theory to be

$$\binom{n + r - 1}{r},$$

e.g., *Feller* [1950, p. 52]. Consequently, the total number of possibilities to be considered for a network with stream numbers $\mathfrak{N}_\Omega$ is

$$F(\mathfrak{N}_\Omega) = \prod_{\omega=1}^{\Omega-1} \binom{N_\omega - 2}{N_\omega - 2N_{\omega+1}} \qquad (7)$$

Note that equation 7 differs by factors of 2^{R_ω} from *Shreve*'s [1966] equation 4 for the number of topologically distinct networks with stream numbers $\mathfrak{N}_\Omega$. These factors allow for the possibility that a tributary may join from either the right or the left; they are important in determining the number of distinct networks but have no effect on the link distribution.

To obtain an expression for $f(\nu; \omega, \mathfrak{N}_\Omega)$, we assume, as required by Shreve's model, that the probability of finding a particular link distribution is just proportional to the number of different ways in which it can occur. A derivation of $f(\nu; \omega, \mathfrak{N}_\Omega)$ for general ω and $\mathfrak{N}_\Omega$ leads to rather complicated algebraic formulas, and it is perhaps more instructive to begin by making an explicit calculation for a third-order basin. In this case, there are $R_1 = N_1 - 2N_2$ first-order streams to be attached to $2N_2 - 1$ links. N_2 of these links form N_2 second-order streams of one link each, and the remaining $N_2 - 1$ make up the single third-order stream. Suppose that we assign k tributaries to the second-order links and $(R_1 - k)$ tributaries to the third-order links. As each tributary added creates one new link, we then have $\nu_2 = N_2 + k$. and $\nu_3 = N_2 - 1 + R_1 - k = N_1 - N_2 - 1 - k$. The number of ways of making this assignment is

$$\binom{N_2 + k - 1}{k}\binom{N_2 - 1 + R_1 - k - 1}{R - k}$$

$$= \binom{N_2 - 1 + k}{k}\binom{N_1 - N_2 - 2 - k}{N_1 - 2N_2 - k}$$

Then

$$f(N_2 + k; 2, \mathfrak{N}_\Omega) = f(N_1 - N_2 - 1 - k; 3, \mathfrak{N}_\Omega) = \binom{N_2 - 1 + k}{k} \cdot \binom{N_1 - N_2 - 2 - k}{N_1 - 2N_2 - k} \Big/ \binom{N_1 - 2}{N_1 - 2N_2}$$

$$k = 0, 1, 2, \cdots, R_1 \qquad (8)$$

since $F(N_1, N_2, 1) = \binom{N_1 - 2}{N_1 - 2N_2}$

We can show by direct summation that the probabilities f are properly normalized. By combining equations 6 and 8, and after some manipulation of combinatorial formulas we find

$$\bar{k} = \frac{N_2(N_1 - 2N_2)}{(2N_2 - 1)} \qquad (9)$$

$$E(\mathfrak{L}; 2, \mathfrak{N}_\Omega) = \frac{N_2(N_1 - 1)}{2N_2 - 1}\,\bar{l}_i \qquad (10)$$

and

$$E(\mathfrak{L}; 3, \mathfrak{N}_\Omega) = \frac{(N_2 - 1)(N_1 - 1)}{2N_2 - 1}\,\bar{l}_i \qquad (11)$$

Note that the distribution in length between second and third orders is made in proportion to the numbers of second- and third-order links in the framework to which the first-order tributaries were attached. Now let

$$E(h; w, h_\Omega) \qquad (12)$$

As mentioned previously, the length parameter most frequently tabulated in the hydrologic literature is $\bar{L}_\omega$, the mean length of streams of order ω. Since $\bar{L}_\omega = \mathfrak{L}_\omega / N_\omega$, the quantity $\bar{L}_\omega$ is in fact just the mean value of $\bar{L}_\omega$ for a collection of networks with stream numbers $\mathfrak{N}_\Omega$.

The results for third-order networks can be generalized to higher order without making a detailed calculation. First, we note that the expressions for $E(\mathfrak{L}; 2, N_\Omega)$ and $\bar{L}_2$ are correct for any value of Ω. Second-order streams can have only first-order streams as tributaries, and equation 10 results from considering the number of ways in which these first-order streams can be attached to N_2 links of second order and $N_2 - 1$ links of higher order. It makes no difference to the results for $E(\mathfrak{L}; 2, N_\Omega)$ and $\bar{\bar{L}}_2$ whether the $N_2 - 1$ links are all third order, as in our example, or whether they are a mixture of third and higher orders, as would be the case for $\Omega > 3$. In a similar way, we can see that, for $\Omega > 3$, $E(\bar{\mathfrak{L}}; 3, N_\Omega)$ will acquire a factor $N_3 (N_2 - 1)/(2N_2 - 1)$ from the assignment of the R_2 second-order tributaries and a further factor of $(N_1 - 1)/(2N_2 - 1)$ from the assignment of the R_1 first-order tributaries. The general result can be written most simply in terms of $\bar{\bar{L}}_\omega$.

$$\bar{\bar{L}}_\omega/\bar{l}_i = \prod_{\alpha=2}^{\omega} (N_{\alpha-1} - 1)/(2N_\alpha - 1) \qquad \omega \geq 2 \qquad (13a)$$

We have not previously specified how $\bar{l}_i$ is to be obtained, but normally the most convenient procedure will be to set it equal to its point estimator, the mean length of the links in the actual network being considered. This definition will be adopted for the remainder of the discussion, without change in notation.

An approximate description of stream length behavior can be obtained by assuming that $\bar{L}_\omega$ can be adequately predicted by using its mean value, $\bar{\bar{L}}_\omega$, i.e.

$$\bar{L}_\omega \approx \bar{\bar{L}}_\omega \qquad (13b)$$

Equation 13b may be regarded as a substitute for Horton's law of stream lengths. In Table 2, the two relations are contrasted for fourth-order basins.

TABLE 2. Mean Stream Length Ratios for Fourth-Order Basins

ω	Horton's law	Equation (13b)
2	$\dfrac{\bar{L}_2}{\bar{L}_1} \approx R_L$	$\dfrac{\bar{L}_2}{\bar{l}_i} \approx \dfrac{N_1 - 1}{2N_2 - 1}$
3	$\dfrac{\bar{L}_3}{\bar{L}_1} \approx R_L^2$	$\dfrac{\bar{L}_3}{\bar{l}_i} \approx \dfrac{(N_1 - 1)(N_2 - 1)}{(2N_2 - 1)(2N_3 - 1)}$
4	$\dfrac{\bar{L}_4}{\bar{L}_1} \approx R_L^3$	$\dfrac{\bar{L}_4}{\bar{l}_i} \approx \dfrac{(N_1 - 1)(N_2 - 1)(N_3 - 1)}{(2N_2 - 1)(2N_3 - 1)}$

Horton's law gives the ratio of $\bar{L}_\omega$ to the mean length of exterior links, whereas equation 13*b* gives the ratio of $\bar{L}_\omega$ to the mean length of interior links. In Horton's law the factor on the right hand increases in simple geometric proportion, whereas in our relation the rate of increase varies with ω. Most important, however, Horton's law merely provides an empirical fit to the data, whereas equation 13*b* gives the mean length ratios in terms of the stream numbers and does not contain any adjustable parameters.

Shreve [1967] has shown that for an infinite topologically random network the bifurcation ratio R_B is 4. He also noted that if all links have the same length (a distribution that satisfies the general conditions discussed above), the stream length ratio is 2. Our results are consistent with Shreve's in the sense that, as the $N_\omega \rightarrow \infty$ and $N_\omega/N_{\omega+1} \rightarrow 4$, equation 13*b* approaches Horton's law with $R_L = 2$.

Broscoe [1959] and *Bowden and Wallis* [1964] have suggested that Horton's law of stream lengths is not valid when Strahler ordering is used. One of the most common sources of deviations from Horton's law is an abnormal length of the single stream of order Ω. We note that equation 13b takes some account of this effect. If L_Ω is unusually small, then it is very likely that $N_{\Omega-1} = 2$, in which case equation 13*b* predicts $L_\Omega = \bar{L}_{\Omega-1}$. On the other hand, if L_Ω is unusually large, then $N_{\Omega-1}$ will probably be considerably greater than 2, and the predicted value of $L_\Omega/\bar{L}_{\Omega-1}$ will also be unusually large. Figure 5 compares *Melton*'s [1957] stream length data on Hog Hollow, Utah (an example cited by Bowden and Wallis) with results obtained from equation 13b and from Horton's law. Equation 13b clearly gives the better representation of the general trend of the data.

(Incidentally, the stream numbers for Hog Hollow ($N_1 = 131$, $N_2 = 22$, $N_3 = 4$, $N_4 = 2$, $N_5 = 1$) are a highly improbable set; when stream numbers for $N_1 = 131$ are ranked in order of probability, this set lies well down in the one percentile group. In general, stream numbers of very low probability tend to produce stream lengths that deviate widely from Horton's law.)

For a more comprehensive test, it would be desirable to have data on groups of basins with the same set of stream numbers. However, stream number data do not exist in such profusion, and we must instead combine results for basins with

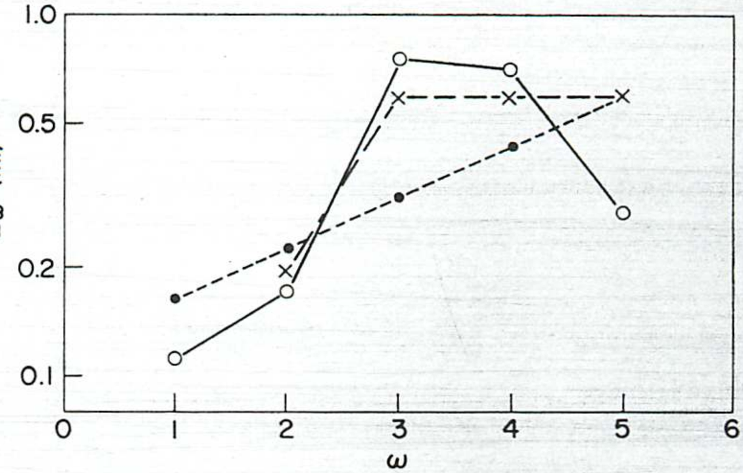

Fig. 5. Mean length versus order for Hog Hollow, Utah. ○ - - - ○; Observation. ● - - - ●; Best fit obtainable with Horton's law of stream lengths. x - - - x; equation 13*b*.

different sets of stream numbers. Figure 6 shows the predicted and observed values of $\bar{L}_2/\bar{l}_i$ and $\bar{L}_3/\bar{l}_i$ for 81 third-order basins investigated by *Melton* [1957]. The relatively good agreement is particularly satisfying in view of the fact that about one-fourth of the basins have ten or fewer sources. Column 4 of Table 3 gives the correlation coefficient of predicted and observed values of $\bar{L}_\omega/\bar{l}_i$ for all of the data we have been able to find. It seems reasonable to infer that our model accounts for about two-thirds of the observed variance in $\bar{L}_\omega/\bar{l}_i$.

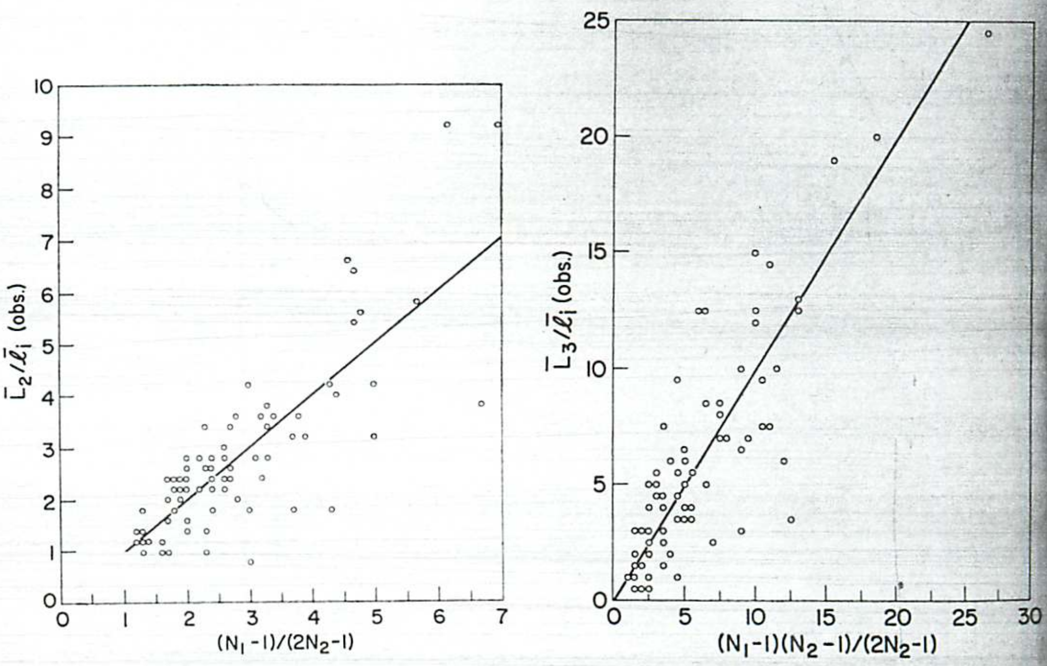

Fig. 6. Left, Observed versus predicted values (equation 13*b*) of $\bar{L}_2/\bar{l}_i$ for 81 third-order basins [*Melton*, 1957]. Right, Same plot for $\bar{L}_3/\bar{l}_i$. In both graphs, the straight lines are not regression lines but are the theoretical curves with no adjustable parameters.

TABLE 3. Correlation of Predicted and Observed Values of $\bar{L}_\omega/\bar{l}_i$

Investigator	Ω	ω	Number of Basins	r Stream Numbers Only	Number of Basins	r Tributary Numbers
Melton [1957]	3	2	81	0.819	44	0.938
		3		0.850		0.944
	4	2	46	0.802	11	0.889
		3		0.826		0.936
		4		0.781		0.933
Coates [1958]	3	2	60	0.687	57	0.880
		3		0.895		0.928

The principal advantages of using equation 13*b* instead of Horton's law are: (1) it gives better insight into the way drainage basins are organized; (2) it eliminates one empirical parameter; (3) it gives much better quantitative predictions for such eccentric cases as the Hog Hollow example; (4) it is applicable regardless of whether or not the basin is complete; (5) finally, the assumptions and approximations involved in obtaining equation 13*b* are known, so that it is clear how the model could be improved and refined. One attempt at improvement is described in the next section.

The above results could be extended in a number of directions. First, detailed properties of the distribution function for L_ω could be calculated using a specific model such as the negative exponential model discussed in a later section. Also, the link distribution function for a single stream can be derived by an approach paralled to the one used above, beginning with the consideration of the number of different ways of assigning k tributaries to one stream and the remaining $R_\omega - k$ tributaries to other streams. The procedures are quite straightforward, but as the data on actual stream systems are not sufficient to make a sensible comparison between theory and observation, we have not carried them through here.

TRIBUTARY NUMBERS

The difference between predicted and observed values of $L_\omega/\bar{l}_i$ may be attributed in part to variation in the number of links per order, in part to variation in link length, and in part to inadequacies of the model. It is worth noting that the deviations due to variation in ν_ω can be completely eliminated by introducing a set of 'tributary numbers' in addition to the usual stream numbers. Let $N_{\alpha\beta}$ be the number of excess tributaries of order α that terminate in a stream of order $\beta > \alpha$. In a network of order Ω, there are $\frac{1}{2}\Omega(\Omega - 1)$ such numbers, but only $\frac{1}{2}(\Omega - 1)(\Omega - 2)$ are independent, since the relation

$$\sum_{\beta=\alpha+1}^{\Omega} N_{\alpha\beta} = R_\alpha$$

$$\alpha = 1, 2, \cdots, \Omega - 1 \qquad (14)$$

must be satisfied. When the $N_{\alpha\beta}$ are known, the numbers of links for each order are completely specified.

$$\nu_\omega = N_\omega + \sum_{\alpha=1}^{\omega-1} N_{\alpha\omega} \qquad (15)$$

Following the procedures of the previous section, we obtain

$$\bar{L}_\omega/\bar{l}_i \approx 1 + \sum_{\alpha=1}^{\omega-1} N_{\alpha\omega}/N_\omega \qquad (16)$$

None of the existing tabulations of data on actual networks gives values of $N_{\alpha\beta}$, but they can of course be determined if outlines of the channel patterns are provided. Both *Melton* [1957] and *Coates* [1958] give the channel patterns, but we were not able to make an unambiguous determination of the tributary numbers in every case. Figure 7 shows observed and predicted values of $\bar{L}_\omega/\bar{l}_i$ for the 44 Melton's third-order networks for which an unambiguous assignment was possible. The last column of Table 3 gives the correlation coefficients between observed and predicted values for all Melton and Coates data. The agreement is considerably improved, and the unexplained variance is reduced to about 15%.

The determination of tributary numbers is of course equivalent to counting the number of links for each order; whether or not the better agreement obtained is worth the extra

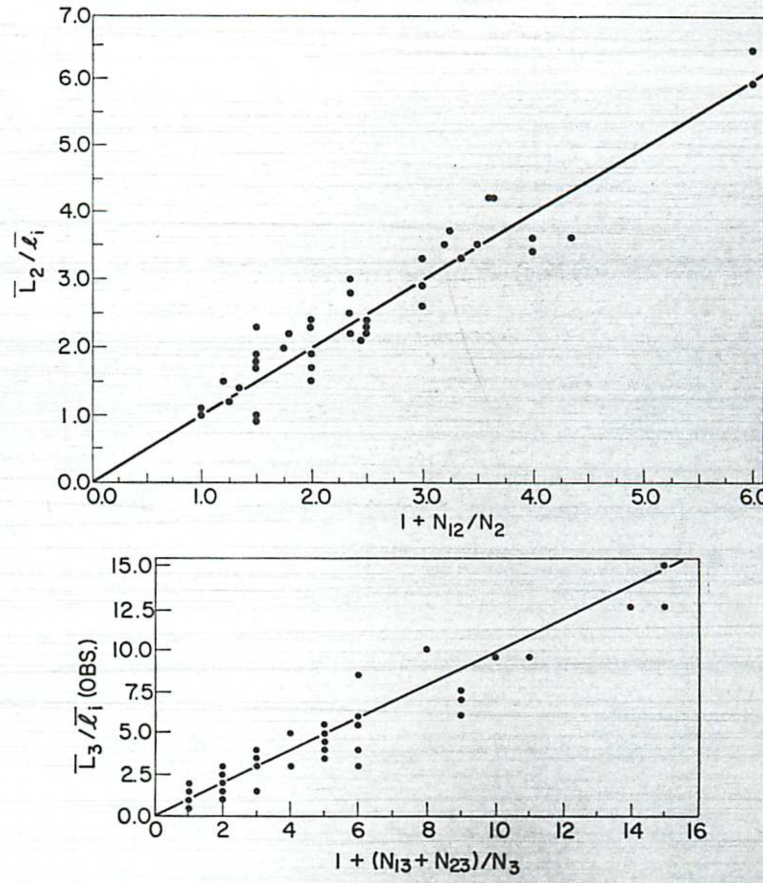

Fig. 7. Same as Figure 6, but with predicted values given by equation 16.

labor involved depends on the needs of the individual worker.

DISTRIBUTION OF LINK LENGTHS

The considerable reduction in variance obtained by using tributary numbers is encouraging and suggests that detailed properties of the distribution of stream lengths could be deduced if our original general assumption about link lengths were replaced with a specific model. One might hope that such a model could be inferred from measurements on actual systems, but, unfortunately, a search of the literature on drainage basin morphology does not provide much help. *Schumm* [1956] and *Maxwell* [1960] have investigated the distribution of first-order stream lengths for a few basins and have concluded that their data could be represented by a log-normal probability density. No results at all have been reported for interior links.

Another possible method of obtaining a model for link length distributions is by simulation techniques. *Leopold and Langbein* [1962] suggested a procedure for simulating channel networks by playing random walk games on a square grid. This scheme is quite laborious when carried out by hand, but *Schenck* [1963] and *Smart et al.* [1968] showed that it can be performed with a digital computer. The latter authors made a detailed study of the properties of such simulated channel networks and, in particular, demonstrated that the lengths of first-order streams have a geometric probability distribution. That is, the probability that a first-order stream has length $L_1 = k + 1$ is

$$P\{L_1 = k + 1\} = p(1 - p)^k \qquad k = 0, 1, 2, \cdots \tag{17}$$

where p is a constant that depends on the exact rules of the random walk game. The most obvious interpretation of this result is that p is the probability that the stream will make a junction (and thus terminate) on any given move, whereas $(1 - p)$ is the probability that it will not make a junction. Thus, equation 17 gives the probability that the last step, which must be a junction, is preceded by k steps without a junction; this is of course just the condition that $L_1 = k + 1$. In recent work we have found that the geometric distribution is also appropriate for interior link lengths in random walk simulated networks.

An important feature of the geometric distribution is that p is independent of the number of steps already made without a junction. We shall assume that this behavior is carried over from simulated systems, where length is a discrete variable, to actual systems, where length is a continuous variable. Then in stream networks, the probability of a junction's occurring in an element of length Δl is independent of the distance already traversed without a junction. Moreover, for sufficiently small elements Δl, the probability of a transition should be proportional to the magnitude of Δl. Thus

$$\lim_{\Delta l \to 0} p = \lambda \, \Delta l \qquad (18)$$

This general type of behavior, in which the probability of an event's occurring in a given interval is constant and independent of previous such events, appears quite commonly in problems involving stochastic processes. In most cases, e.g., radioactive decay or queuing theory, the independent variable is time. A large body of literature exists on these subjects, and we can adapt a number of its results to our purposes simply by replacing time by length [*Feller*, 1950, pp. 218–221].

For example, if we start at some arbitrary point in a channel network and traverse a distance l, the probability of finding exactly j junctions is the Poisson function

$$p(j; \lambda l) = e^{-\lambda l}(\lambda l)^j / j! \qquad (19)$$

In particular, the probability of finding no junction at all is

$$p(0; l\lambda) = e^{-\lambda l}$$

The mean and variance of j are both equal to $l\lambda$.

$$E(j) = \sum_{j=0}^{\infty} jp(j; \lambda l) = \lambda l \qquad (20)$$

$$V(j) = E(j^2) - E^2(j) = \lambda l \qquad (21)$$

These results should be useful in various practical considerations about drainage basins, such as the optimum placement of gaging stations.

For the stream length problem, we are especially interested in knowing the probability $P(L;\ n, \lambda)dL$ that n consecutive links in a channel network have a total length lying between L and $L + dL$.

$$\begin{aligned} P(L; n, \lambda)\, dL &= p(n - 1; \lambda L)\lambda \, dL \\ &= e^{-\lambda L}[(\lambda L)^{n-1}/(n - 1)!]\lambda \, dL \qquad (22) \end{aligned}$$

where $p(n - 1;\ \lambda L)$ is the probability that $n - 1$ junctions occur in the distance L, and λdL is the probability that the nth junction occurs in the element dL. $P(L;\ n, \lambda)$ is just the well known gamma density (perhaps more usually expressed as a function of the parameter $\beta = 1/\lambda$)

$$E(L) = \int_0^\infty Lp(n - 1; \lambda L)\lambda \, dL = n/\lambda \qquad (23)$$

$$V(L) = E(L^2) - E^2(L) = n/\lambda^2 \qquad (24)$$

For the important case $n = 1$, $P(L, 1, \lambda) = \lambda\ e^{-\lambda L}$, the negative exponential density.

As mentioned previously, there are no published data on length distributions for interior links. To have at least some comparison with observation, we have made measurements of link lengths in Gourd Creek and Coalpit Hollow, two contiguous small watersheds in Missouri. The U. S. Geological Survey 1:24,000 map (Yancy Mills quadrangle) was enlarged by a factor of 2, and special care was taken in drying the prints to keep the distortion caused by shrinkage at 1% or less. Lengths were measured with a Dietzgen 1719B Map Measurer and were generally reproducible to 1/32 inch. Thus, the measurement errors are probably less than those due to inaccuracies in the maps.

Both exterior and interior links in the Gourd Creek-Coalpit Hollow area had length distributions that were highly right-skewed and at least approximately negative exponential. How-

TABLE 4. Frequency Distribution of Interior Link Lengths, Gourd Creek and Coalpit Hollow

Range of L (mi)	Observed Frequency	Expected Frequency
0.0–0.05	39	44.62
0.05–0.10	36	30.40
0.10–0.15	23	20.71
0.15–0.20	17	14.11
0.20–0.25	7	9.61
0.25–0.30	10	6.55
>0.30	8	14.00

$\bar{L} = 0.130$ mi $\sigma_L{}^2 = 0.0158$ mi^2
$-2\ ln\ \lambda = 7.91$ $\chi_{0.05}{}^2(5) = 11.07$

ever, the two distributions differed substantially. The determination of exterior link lengths is clearly a much more subjective procedure than is the determination of interior link lengths, especially when no field measurements are made. Consequently, only the interior link data have been compared in detail with the model. Table 4 lists the frequencies of 140 interior link lengths for intervals of 0.05 mile. If, as required by our model, the probability density is negative exponential, then the probability that a link has a length less than L' is $(1 - e^{-\lambda L'})$. The point estimator of λ for the gamma density is $n/\bar{L}$ and for the data of Table 4 is 7.675 mi^{-1}. Theoretical frequencies calculated with this value of λ are given in the third column of the table. The goodness-of-fit test gives 7.91 for $-2\ ln\ \lambda$, to be compared with 11.07 for χ^2 with 5 degre of freedom and 5% significance. (Here λ is th goodness-of-fit parameter, not 7.675.) Figure compares the observed and calculated cumu lative frequency distributions. The agreeme between observation and the proposed mod although far from perfect, is at least sufficient good to justify using the model as a basis f discussion of stream length statistics.

To make sure that the above results we not peculiar to Gourd Creek, supplementa measurements of a lower degree of precisi were made on two other drainage basins, o in the Piedmont province of Virginia and o in the Basin and Range province of Arizo In both cases, the results obtained were qua tatively similar to those for Gourd Creek; t goodness of fit was about the same, and t most obvious disagreement between observ and theoretical values was that the observ distribution did not have enough very sho links. It is relevant to note that, for all thr networks, the agreement with the model cou be improved appreciably by replacing L i equation 22 with $(L - L_0)$. As an example, i for the Gourd Creek data we rather arbitrarily choose $L_0 = 0.01$ mi, the value of $-2\ ln\lambda$ is reduced from 7.91 to 3.07. This modification of the distribution function is equivalent to assuming that the minimum length for interior links is not zero but L_0. Such an assumption is not unreasonable, for various physical considerations suggest that very short links may

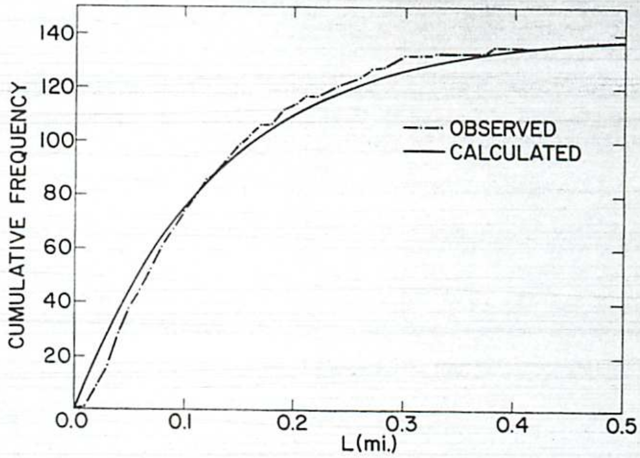

Fig. 8. Comparison of observed and calculated cumulative frequencies of interior link lengths for Gourd Creek and Coalpit Hollow, Mo.

e unstable. It appears, then, that a useful, emiempirical distribution function for interior ink lengths would be the negative exponential istribution modified so as to have at least ome bias against very short links.

MEAN AREA RATIOS

The procedures of the second section can, ith two additional assumptions, be extended area relationships. An expression for the rainage density is obviously required and, nce basins with $\omega \geq 2$ contain first-order basins, e also need to know the distribution of lengths r first-order streams. We assume that (1) The rainage density is uniform, i.e., the area drained y a first-order stream of length L_1 is L_1/D, nd the interbasin area for an interior link of ngth l is l/D, where D is the drainage density. 2) Lengths of both interior and exterior links re independent random variables drawn from e same population. The first approximation generally, although by no means always, tisfactory for basins which are not too large. he second approximation is probably often ther bad and is made here only to help to duce the considerable amount of algebra in- olved. Even with this simplification we could ot obtain a general expression for $\bar{A}_\omega/\bar{A}_1$, and e individual expressions become more and ore unwieldy as ω increases. For $\omega = 2$ and 3, find, in analogy with equation 13*b*

$$\approx \frac{2N_1 + 2N_2 - 3}{2N_2 - 1} \qquad (25)$$

$$\approx \frac{4(N_1N_2 + N_2N_3 + N_1N_3)}{(2N_2 - 1)(2N_3 - 1)} - \frac{6(N_1 + N_2 + N_3) + 7}{(2N_2 - 1)(2N_3 - 1)} \qquad (26)$$

n the limit of an infinite topologically random work, $\bar{A}_2/\bar{A}_1 \rightarrow 5$ and $\bar{A}_3/\bar{A}_1 \rightarrow 21$. *Shreve* 67] suggested that in this limit the area re- ons reduced to Horton's law with $R_A = 4$. wever, it appears that our results are con- ent with the remainder of Shreve's paper. very large random networks, $N_\omega \rightarrow N_1/4^{(\omega-1)}$ $\nu_\omega \rightarrow N_\omega 2^{(\omega-1)}$. Building up a network ording to these rules leads directly to the uence 1, 5, 21, 85, 341, $\cdots$, $4(\bar{A}_{\omega-1}/\bar{A}_1) + 1$, for $\bar{A}_\omega/\bar{A}_1$. Probably the best existing ap- ximation to infinite topologically random networks with uniform drainage density is the *R*1 fifth-order random walk networks discussed by *Smart et al.* [1968]. For these systems N_1 ranges between 63 and 213, and the mean areas drained by interior and exterior links are 1.99 and 2.02 units, respectively; see the original paper for further details. For the sample of 24 networks, the median values of $\bar{A}_2/\bar{A}_1$ and $\bar{A}_3/\bar{A}_1$ are 5.13 and 22.1, respectively.

Because data on mean basin areas are very scarce, no attempt has been made to compare equations 25 and 26 with observed values of $\bar{A}_2/\bar{A}_1$ and $\bar{A}_3/\bar{A}_1$. However, it seems likely that the quantitative agreement would be somewhat worse than in the stream length case because of the extra assumptions involved.

Acknowledgment. I am indebted to D. V. Judd for preparing the accurate enlargements of the Gourd Creek map.

REFERENCES

Bowden, K. L., and J. R. Wallis, Effect of stream-ordering technique on Horton's laws of drainage composition, *Geol. Soc. Am. Bull. 75*, 767–774, 1964.

Broscoe, A. J., Quantitative analysis of longitudinal stream profiles of small watersheds, *Proj. NR 389-042 Tech. Rept. 18*, Dept. Geol., Columbia Univ., 1959.

Coates, D. R., Quantitative geomorophology of small drainage basins of southern Indiana, *Proj. NR 389-042 Tech. Rept. 10*, Dept. of Geol., Columbia Univ., 1958.

Feller, W., *Probability Theory and Its Applications*, John Wiley & Sons, New York, 1950.

Horton, R. E., Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative geomorphology, *Bull. Geol. Soc. Am., 56*, 275–370, 1945.

Leopold, L. B., and W. B. Langbein, The concept of entropy in landscape evolution, *U. S. Geol. Surv. Profess. Paper 500-A*, 1962.

Leopold, L. B., and J. P. Miller, Ephemeral streams: Hydraulic factors and their relation to the drainage net, *U. S. Geol. Surv. Profess. Paper 282-A*, 1956.

Maxwell, J. C., Quantitative geomorphclgy of the San Dimas National Forest, *Proj. NR 389-042, Tech. Rept. 19, Dept. Geol.*, Columbia Univ., 1960.

Melton, M. A., An analysis of the relations among elements of climate, surface properties, and geomorphology, *Proj. NR 389-042 Tech. Rept. 11*, Dept. Geol., Columbia Univ., 1957.

Morisawa, M. E., Quantitative geomorphology of some watersheds in the Appalachian Plateau, *Bull. Geol. Soc. Am., 73*, 1025–1046, 1962.

Scheidegger, A. E., The algebra of stream order numbers, *U. S. Geol. Surv. Profess. Paper 525-B,* 1965.

Schenck, H. S., Jr., Simulation of the evolution of drainage-basin networks with a digital computer, *J. Geophys. Res., 68,* 5739–5745, 1963.

Schumm, S. A., Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey, *Bull. Geol. Soc. Am., 67,* 597–646, 1956.

Shreve, R. L., Statistical law of stream numbers, *J. Geol., 74,* 17–37, 1966.

Shreve, R. L., Infinite topologically random channel networks, *J. Geol., 75,* 178–186, 1967.

Smart, J. S., A. J. Surkan, and J. P. Considine, Digital simulation of channel networks, *IASH Proc. 14th IUGG Assembly* (in press), 1968.

Strahler, A. N., Hypsometric analysis of erosional topography, *Bull. Geol. Soc. Am., 63,* 1117–114[illegible], 1952.

(Manuscript received May 3, 1968.)