STATISTICS OF HYDROLOGIC NETWORKS ON FRACTIONAL BROWNIAN SURFACES

Michael F. Goodchild
Brian Klinkenberg
Maria Glieca
Mahbub Hasan
Department of Geography
The University of Western Ontario
London, Ontario, Canada N6A 5C2

ABSTRACT

The fractional Brownian surfaces described by Mandelbrot (1) in connection with terrain simulation can be used to investigate the basis for morphometric laws. Empirical regularities of geomorphology which can be obtained equally from simulated surfaces where no geomorphic processes have operated should be interpreted as statistical artifacts. There are certain ways in which real drainage networks are observed to deviate from the predictions of the random topology model of Shreve (2,3). Analysis of drainage networks on simulated surfaces shows similar deviations.

INTRODUCTION

The notion of using fractional Brownian (fBm) surfaces for the simulation of terrain originates with the work of Mandelbrot and has attracted considerable attention (for review see (1)). Goodchild (4) compared an island off the coast of Newfoundland to the fBm model, and Mark and Aronson (5) have examined the fit of the model to terrain in the eastern US. These fBm surfaces differ from real terrain in many ways, one of the more obvious ones being the equal abundance of peaks and pits, which is suggestive perhaps of karst or dead-ice topographies but not of more general eroded landscapes. Nevertheless the fractional Brownian process has certain attractions when compared to other available methods of simulation. The self-similarity property ensures the absence of scale dependence in the simulated landscape, so that it is impossible to distinguish statistically between a part of the landscape at one scale, and a smaller piece displayed at a larger scale so as to appear of the same size. Since most if not all geomorphic processes are scale-dependent, having different degrees of effect on the landscape at different scales, a self-similar or scale-independent simulation can be thought of as free of geomorphological influence. fBm surfaces are thus a useful starting point for the simulation of erosion processes. Furthermore, this raises the possibility of using fBm surfaces as norms or standards against which real, eroded terrain can be compared. Empirical laws which are found to be true of real terrains but not of the simulated ones can therefore be interpreted as the result of the operation of geomorphic processes, whereas any empirical regularities which are found to be true of both real and simulated terrains are by implication statistical artifacts of no geomorphological significance. Early simulations of drainage basin development by Schenck (6) and others showed that randomly joining streams obeyed similar laws to those found by Horton (7) for real stream networks, implying that it is not adherence to but deviation from these laws that is significant in a geologic or geomorphologic sense.

The work of Shreve (2,3) placed these observations on a firm mathematical basis and led to the random model of stream network development. All networks are assumed to be part of an infinite topologically random channel network (TRCN) in which streams combine randomly so that all possible topologies are equally likely. Later developments of the model by Shreve and others led to predictions about associated geometries, but in this paper we are concerned only with the notion of random topology and the results which derive from it.

Empirical tests of the random topology model have recently been reviewed by Abrahams (8), who notes several ways in which real networks are consistently observed to differ from the random model. For large basins (number of first order streams $M > 50$) there is a tendency for bifurcation ratios to exceed the model's predictions, particularly in areas of high relief. Similarly the proportion of first order streams which are TS links (join to streams of order greater than 1) is found to be greater than the theoretical value of 50%. Abrahams (9) found a strong correlation between the amount of deviation from the model and the relative relief of the topography, whereas Howard (10) proposed a mechanism based on stream capture.

The method of simulation used by Schenk (6) and others was to allow streams to wander randomly over a square lattice, generating a move at each lattice cell from a random integer corresponding to one of four possible move directions. Special means had to be devised to avoid topological inconsistencies such as closed loops and spirals. The method used by Seginer (11) avoided these problems by first generating a random surface, and then extracting the stream network from it, thus ensuring topological consistency. A study by Craig (12) of simulated erosion processes on surfaces extracted stream networks and compared them to the random topology model as a means of validation of the simulation process, a dubious procedure given the known departures of real networks from the model.

Although both of these methods of random network generation are found to yield results in broad agreement with the random topology model, they clearly differ from it in principle. The requirement that adjacent basins pack together on the lattice ensures that basins are not statistically independent, and therefore that all topologies are not equally likely. The use of a topographic surface to obtain the stream networks imposes additional constraints of smoothness and continuity of elevation. Although Seginer obtained his surfaces by generating independent elevations in each cell, both real and fBm surfaces are comparatively smooth, ensuring strong interdependence between flow directions in adjacent cells.

The effect of such constraints on the random topology model is unknown, and is unlikely to yield to mathematical analysis. The approach proposed in this paper is to analyze stream networks obtained from fBm surfaces. If broad agreement is found with the random topology model, the conclusion will be that the effects of the constraints are minimal. On the other hand if deviations are found from the random topology model, it will be possible to compare these to the deviations observed in real networks. And if the deviations are consistent, the implication will be that they result from topographical constraints on the operation of the TRCN model on real surfaces, rather than from any geomorphologic mechanism.

SIMULATIONS

Arrays of 258 by 258 elevations were generated using the method described by Mandelbrot (13) in which an initially flat surface is successively distorted by displacement along randomly located 'fault' lines. Mandelbrot (14) has argued that the more commonly used method described by Fournier et al. (15) is statistically incorrect. The profile of the displacement on either side of each 'fault' can be adjusted to produce surfaces with different values of the parameter H, which ranges from 0 to 1. A low H yields a rugged surface with relatively high variance locally; a high H corresponds to a smooth surface with little local variance but with significant drift. H was varied from .3 to .7 in steps of .1, and 1000 'faults' were generated for each surface.

Each cell was then assigned an integer indicating flow direction according to the following rules. First, all border cells were designated sinks (direction 0). Second, if at least one of a cell's four neighbours was strictly lower in elevation, the direction was toward the neighbour of lowest elevation. Finally all other cells were designated sinks. This classification is similar to that used by O'Callaghan and Mark (16) (but compare (17)) except that it is based on four rather than eight neighbours and only four directions are allowed. The resulting frequencies of directions and sinks varied with H (Table 1), particularly on high H surfaces where frequencies also varied markedly with direction because of the inherent smoothness in the surface, although the direction of bias is of course random. Three replications were made of the H=0.7 simulation for comparison and because this value of H was thought by Mandelbrot (13) to give the closest resemblance to real landscapes.

Table 1

Frequencies of Sinks and Flow Directions

H	0 (Sink)	1 (North)	2 (East)	3 (South)	4 (West)
.3	7319	14373	14802	14316	14726
.4	5902	14051	15423	15779	14381
.5	4536	13775	15524	17255	14446
.6	2521	15007	9096	14761	24151
.7a	462	46343	5232	999	12500
.7b	1016	4492	22112	30131	7785
.7c	1199	8396	17193	27430	11318

Next, border cells were discarded and all drainage basins were extracted from the remaining 256 by 256 array for each surface. The algorithm processed each array row by row. On encountering a sink, all drainage networks discharging into that sink were identified in the form of binary strings (3). In our implementation of the binary string concept the basin is traversed from the outlet from left to right. A zero is coded the first time each interior link is traversed (equivalent to coding a zero the first time a junction is reached) and a one is coded each time an exterior link is encountered. No attempt was made to apply a criterion for channel development such as that used by O'Callaghan and Mark (16); exactly one channel was assumed to flow from every cell in the direction inferred from the cell's slope. Because a square array was used, there were frequent instances of four-valent junctions, whereas all theoretical work on lake-free channel networks has assumed that all junctions are three-valent; this problem was resolved by creating two three-valent junctions, the incoming stream immediately clockwise of the outgoing stream being assumed to join downstream of the other two. In terms of the binary coding scheme, which proceeds from left to right around the tree, this process of breaking four-valent junctions always yields ...010... rather than ...00...1...

All analyses of the basins obtained in this way were carried out by processing the binary strings. The streams in a basin can be classified according to the Strahler ordering system by replacing all instances of 0xx by x+1, and all instances of 0xy, x not equal to y, by the greater of x and y. Each time the first of these conditions occurs the counter for streams of order x+1 is incremented. This process must terminate with a single digit equal to the order of the basin.

ANALYSIS

In the infinite TRCN model the probability that a randomly chosen basin or subbasin is of order w is $1/2^w$. The relative abundances of basins on the fBm surfaces are shown in Table 2.

Table 2

Abundances of Basins by Order

H	1	2	3	4	5	Total
.3	14107	4948	648	1	0	19704
.4	10596	4512	794	5	0	15907
.5	7474	3590	822	15	0	11901
.6	3902	2139	743	40	1	6825
.7a	622	379	180	51	8	1240
.7b	1543	801	391	55	4	2794
.7c	1796	998	466	70	0	3330

The model predicts the number of order w+1 basins to be half the number of order w basins, and this is clearly most nearly true of the lower orders of the high-H surfaces. All surfaces show severe truncation of the high orders. The very rugged topography and high incidence of sinks on the low H surfaces have clearly led to an overabundance of first order basins.

The probability that a randomly chosen stream in an infinite TRCN has order w is $3/4^w$. Again we find better fit to the model on the smoother surfaces of high H, and an overabundance of low-order streams particularly on low H surfaces. To fit the model the number of streams of order w+1 in the following table should be one quarter of the number of streams of order w, for each surface.

Table 3

Abundances of Streams by Order

H	1	2	3	4	5	Total
.3	32016	6318	650	1	0	38985
.4	29038	6264	804	5	0	36111
.5	25365	5512	852	15	0	31744
.6	19888	4214	833	42	1	24978
.7a	9086	1583	365	67	8	11109
.7b	13489	2481	556	65	4	16595
.7c	14168	2707	624	70	0	17569

Analysis of full basin topologies is limited by the explosive growth in the number of possible distinct topologies as the magnitude of the basin (number of first order streams M) increases. In this study we have limited the analysis to basins of magnitude 5, for which there are 14 possible topologies. For convenience we will refer to each basin by the octal representation of its binary string: thus for example the basin represented by the string 0101011, which is one of the five topologies for magnitude 4 basins, will be referred to as basin 53. The four others are 17, 27, 33 and 47. Basins can be counted very quickly by using this octal representation (or its decimal equivalent) as a hash code.

The abundances of the five possible topologies for magnitude 4 basins are shown in Table 4 below.

Table 4

Abundances of Topologies for Magnitude 4 Basins

H	17	27	33	47	53	Total
.3	38	72	165	211	245	731
.4	60	96	169	218	209	752
.5	71	88	112	163	162	596
.6	46	77	66	86	84	359
.7a	7	13	7	11	17	55
.7b	25	26	25	30	36	142
.7c	21	29	29	34	50	163

In the infinite TRCN model the five topologies are equiprobable, but this is clearly not true of the simulations. Unfortunately the method of breaking four-valent junctions creates a consistent bias. Three first-order streams meeting at a junction are always recoded to 01011 rather than 00111, and hence of the two possible topologies for magnitude 3 basins, 13 is observed much more frequently than 7. In the case of magnitude 4 basins the bias favours 53, 47 and 27 over 17, as is evident in the table above.

We can remove the effect of bias by randomizing the splitting of four-valent junctions, or by aggregating topologies to ambilateral classes. Two topologies are said to be of the same ambilateral class if one can be obtained from the other by a process of switching the two incoming subbasins at one or more junctions, in other words by ignoring left-right distinctions. Smart (18) first suggested aggregating in this way as a means of reducing the number of possible topologies for high magnitude basins. The algorithm to identify a basin's ambilateral class operates on the binary string representation. Each zero in a valid string represents a junction, and is followed by two blocks of binary code representing the two subbasins incident at that junction. Each block of code is evaluated as an octal (or decimal) number, and the order of the two blocks is reversed if necessary so that the first block has a lower numerical value than the second. For example, consider the string 00111 (basin 7, magnitude 3). The first zero is followed by the blocks 011 and 1, with numerical values 3 and 1 respectively. Since the first is greater than the second, they are reversed to give the string 01011, basin 13, which is the only ambilateral class for magnitude 3. In this way the five topologically distinct magnitude 4 basins recode to two ambilateral classes, 0101011, basin 53, and 0011011, basin 33. The relative abundances of ambilateral classes should not be affected by the method used to break up four-valent junctions.

Although Werner and Smart (19) give recursive formulae for the number of ambilateral classes for each magnitude, the number of topologies corresponding to each must be enumerated by inspection. Table 5 gives the observed and expected abundances for magnitudes 4, 5, 6 and 7 basins for surface H=.3. The results for the other surfaces are similar.

The Null Hypothesis represented by the infinite TRCN model is rejected at the .10 level for M=4 and at the .05 level for all other magnitudes for this surface, and similar patterns are observed for the other surfaces. In each case the more probable ambilateral classes tend to be more abundant than expected, and the less probable classes to be less abundant. High relative probabilities occur in ambilateral classes with relatively large numbers of asymmetrical junctions. For example, in the string 0011011 the two blocks following the first zero are identical, so the corresponding ambilateral class has only one member, whereas the string 0101011, which has the same magnitude, has three other members of the same class (0100111, 0010111 and 0001111) because two of its zeroes now represent asymmetrical junctions.

At any symmetrical junction the downstream link is of order one higher than both of the incoming links. Thus in general basins with a large percentage of symmetrical junctions

Table 5

Abundances of Ambilateral Classes, Surface H=.3

Magnitude	Class	Observed	Probability	Expected	p(TS)
4	53	566	4/5	546	2/4
	33	165	1/5	185	0/4
5	253	258	8/14	230	3/5
	153	110	4/14	115	1/5
	233	35	2/14	58	1/5
6	1253	82	16/42	73.5	4/6
	653	39	8/42	36.7	2/6
	1153	23	8/42	36.7	2/6
	553	24	4/42	18.3	2/6
	1233	21	4/42	18.3	2/6
	633	4	2/42	9.1	0/6
7	5253	40	32/132	30.8	5/7
	3253	20	16/132	15.4	3/7
	4653	10	16/132	15.4	3/7
	2653	19	16/132	15.4	3/7
	5153	20	16/132	15.4	3/7
	3153	4	8/132	7.7	1/7
	4553	5	8/132	7.7	3/7
	5233	5	8/132	7.7	3/7
	3233	2	4/132	3.8	1/7
	4633	0	4/132	3.8	1/7
	2633	2	4/132	3.8	1/7

tend to be of higher order, everything else being equal. The table confirms a tendency for basins of high order to have few ambilateral classes and to be less abundant than expected, while basins of low order, <u>ceteris paribus</u>, tend to be overabundant.

All first order streams can be placed into two classes depending on whether they join another first order stream or a stream of higher order (another exterior link or an interior link respectively). The first type are denoted S for source links and the second TS for tributary source links. S links must by definition terminate in symmetrical junctions, so we expect a correlation between the proportion of links in an ambilateral class which are TS and the probability of that class in the TRCN model. The percent TS is shown in Table 5 and confirms an overabundance of basins with high percentages. Furthermore the highest percentages of TS links are found in basins with one long second order stream, with all but two of the first order streams spaced along its length, with a binary representation of the form n(01)011. It would appear likely that such basins would be relatively elongated under a wide range of geometrical assumptions.

To test these hypotheses more directly the percentage of TS links in each basin was tabulated by basin magnitude for each surface. Pairs of S links are readily identified by the occurrence of ..011.. in the binary representation; all other ones are TS links. In addition the bifurcation ratio was estimated for each basin by two methods, the (w-1)th root of the number of first order streams (B0), and the ratio of first order to second order streams (B1). Table 6 shows the associated statistics for surface H=.7c tabulated by basin magnitude up to M=20; again, the other surfaces show similar results.

In Table 6 the values of p(TS), or the probability that a randomly chosen first order stream is TS in the infinite TRCN model, is given by (M-2)/(2M-3) (20). It is easy to show that the limit of this ratio as M tends to infinity must be 1/2, as follows. For an infinite basin any binary digit is statistically independent of its neighbours in the string and is equally likely to be a zero or a one. There are eight equiprobable combinations for the two digits in front of a one and the digit following, four of which dictate that the embedded one be an S link, and the remaining four require a TS link:

0010 - TS   0011 - S   0110 - S   0111 - S
1010 - TS   1011 - S   1110 - TS   1111 - TS

However, Table 6 shows clearly that the percentage of TS links observed is much more than expected.

Table 6

Bifurcation and %TS by Basin Magnitude, Surface H=.7c

M	n	%TS	p(TS)	B0	B1
3	4912	33.3	.333	3.00	3.00
4	2798	39.8	.400	3.59	3.59
5	1686	46.7	.429	4.08	4.17
6	1045	49.7	.444	4.25	4.51
7	753	53.1	.455	4.43	4.87
8	542	52.8	.462	4.21	4.87
9	372	55.9	.467	4.35	5.20
10	259	56.8	.471	4.46	5.38
11	221	58.7	.474	4.49	5.62
12	147	57.9	.476	4.31	5.50
13	149	59.9	.478	4.39	5.75
14	104	60.2	.480	4.79	5.99
15	77	60.0	.482	4.12	5.64
16	83	61.4	.483	4.60	6.05
17	68	63.3	.484	4.03	5.99
18	46	61.8	.485	4.63	6.03
19	48	64.3	.486	4.18	5.98
20	37	66.2	.487	4.38	6.48

CONCLUSIONS

The simulated drainage networks obtained from fBm surfaces differ from the predictions of the infinite TRCN model in several ways, and in many of these there is agreement with previously observed deviations between the random model and real networks. More specifically, and following Abrahams (8, p.164), it is observed that:

1) The percentage of TS links exceeds model predictions, particularly in high-magnitude basins.

2) The excess of TS links increases with magnitude.

3) For both evaluations of the bifurcation ratio, there is a tendency for large basins to exceed the predictions of the model.

4) Bifurcation ratios increase with M.

5) B1 generally exceeds B0.

Abrahams notes further inconsistencies involving asymmetry, but for the reasons noted above the method of obtaining networks from simulated surfaces does not allow for the analysis of asymmetry.

We conclude that many of the ways in which real drainage networks have been observed to differ from the predictions of the random topology model are replicated in these simulations. Because the model does not consider the constraints imposed on network topology by surface smoothness and by the need to pack basins together, it may suffer from basic inadequacies which prevent it from fully describing the appropriate Null Hypothesis for stream networks. As a result many properties of observed networks may be incorrectly interpreted as having geomorphological significance.

It would be unwise to argue at this stage that the infinite TRCN model be replaced by fBm simulations as a norm for geomorphological interpretation. Although the latter reproduce several observed inconsistencies, we have already noted that the presence of abundant pits is incompatible with most real terrain. In addition several of the assumptions made in deriving networks are inconsistent with reality, including the lack of a criterion to distinguish between overland and channel flow. The model clearly needs refinement, and it is possible that such refinement will remove the agreement between its predictions and reality, in other words demonstrating that such agreement is merely coincidental.

It would be possible to deal with the problem of pits by treating them as lakes. Although a lake-rich network can be reduced to a three-valent planted tree, to do so would be to ignore the constraints imposed by the lakes on the possible topologies. Alternatively one could treat the network as lake rich using the ideas developed by Mark and Goodchild (21) and subsequently analyzed by Mark (22) and Mark and Averack (23). Finally various models of erosion could be used to reduce the abundance of lakes (compare for example (12,24)).

REFERENCES

1. Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. Freeman, San Francisco.
2. Shreve, R.L., 1966. Statistical law of stream numbers. Journal of Geology 74 17-37.
3. Shreve, R.L., 1967. Infinite topologically random channel networks. Journal of Geology 75 178-186.
4. Goodchild, M.F., 1982. The fractional Brownian process as a terrain simulation model. Modelling and Simulation (Proceedings of the Thirteenth Pittsburgh Conference) 13 1133-1137.
5. Mark, D.M. and P.B. Aronson, 1985. Scale-dependent fractal dimensions of topographic surfaces. Mathematical Geology (to appear).
6. Schenck, H.S. Jr., 1963. Simulation of the evolution of drainage basin networks with a digital computer. Journal of Geophysical Research 68 5739-5745.
7. Horton, R.E., 1945. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. Bulletin, Geological Society of America 56 275-370.
8. Abrahams, A.D., 1984. Channel networks: a geomorphological perspective. Water Resources Research 20 (2) 161-188.
9. Abrahams, A.D., 1977. The factor of relief in the evolution of channel networks in mature drainage basins. American Journal of Science 277 626-645.
10. Howard, A.D., 1971. Simulation model of stream capture. Bulletin, Geological Society of America 82 1355-1376.
11. Seginer, I., 1969. Random walk and random roughness models of drainage networks. Water Resources Research 5 591-607.
12. Craig, R.G., 1980. A computer program for the simulation of landform erosion. Computers and Geosciences 5 111-142.
13. Mandelbrot, B.B., 1977. Fractals: Form, Chance and Dimension. Freeman, San Francisco.
14. Mandelbrot, B.B., 1982. Comment on computer rendering of fractal stochastic models. Communications, ACM 25 581-583.
15. Fournier, A., D. Fussell and L. Carpenter, 1982. Computer rendering of stochastic models. Communications ACM 25 371-384.
16. O'Callaghan, J.F. and D.M. Mark, 1984. The extraction of drainage networks from digital elevation data. Computer Vision, Graphics and Image Processing 28 323-344.
17. Jenson, S.K., 1985. Automated derivation of hydrologic basin characteristics from digital elevation model data. Proceedings, AutoCarto 7, Washington, 301-310.
18. Smart, J.S., 1969. Topological properties of channel networks. Bulletin, Geological Society of America 80 1757-1774.
19. Werner, C. and J.S. Smart, 1973. Some new methods of topologic classification of channel networks. Geographical Analysis 5 (4) 271-295.
20. Mock, S.J., 1971. A classification of channel links in stream networks. Water Resources Research 7 1558-1566.
21. Mark, D.M. and M.F. Goodchild, 1982. Topologic model for drainage networks with lakes. Water Resources Research 18 275-280.
22. Mark, D.M., 1983. On the composition of drainage networks containing lakes: statistical distribution of lake in-degrees. Geographical Analysis 15 97-106.
23. Mark, D.M. and R. Averack, 1984. Link length distributions in drainage networks with lakes. Water Resources Research 20 (4) 457-462.
24. Hugus, M.K. and D.M. Mark, 1984. Spatial data processing for digital simulation of erosion. Proceedings, Fall ASP-ACSM Convention, San Antonio, 683-693.