Geomorphology, 5 (1992) 59–76
Elsevier Science Publishers B.V., Amsterdam

A physical basis for drainage density

David G. Tarboton[a], Rafael L. Bras[b] and Ignacio Rodriguez-Iturbe[c]
[a]*Utah Water Research Laboratory, Utah State University, Logan, UT, USA*
[b]*Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*
[c]*Instituto Internacional de Estudios Avanzados, Caracas, Venezuela*

(Received April 19, 1991; accepted after revision October 19, 1991)

ABSTRACT

Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1992. A physical basis for drainage density. In: R.S. Snow and L. Mayer (Editors), Fractals in Geomorphology. *Geomorphology*, 5: 59–76.

Drainage density, a basic length scale in the landscape, is recognized to be the transition point between scales where unstable channel-forming processes yield to stable diffusive processes. This notion is examined in terms of equations for the evolution of landscapes that include the minimum necessary mathematical complexity. The equations, a version of the equations studied by Smith and Bretherton (1972), consist of conservation of sediment, an assumption that sediment movement is in the steepest downslope direction, and a constitutive relationship which gives the sediment transport rate as a function of slope and upslope area. The difference between processes is embedded in the constitutive relation. Instability to a small perturbation can be determined according to a criteria given by Smith and Bretherton and results when the sediment transport rate is strongly dependent on upslope area, whereas stability occurs if the main dependence is on slope. Where multiple processes are present, the transition from stability to instability occurs at a particular scale. Based on the idea that instability leads to channelization, the transition scale gives the drainage density. This scale can be determined as a maximum, or turn over point in a slope–area scaling function, and can be used practically to determine drainage density from digital elevation data. Fundamentally different scaling behavior, an example of which is the slope–area scaling, is to be expected in the stable and unstable regimes below and above the basic scale. This could explain the scale-dependent fractal dimension measurements that have been reported by others.

Introduction

In analyzing, characterizing, and trying to understand landscape form, a problem at the heart of geomorphology, the notion of drainage density, defined by Horton (1932, 1945) is fundamental. Drainage density is defined as:

$$D_d = L_T / A \qquad (1)$$

where L_T is the total length of streams and A is contributing area. D_d has inverse length units, and the inverse of D_d gives a length scale associated with the landscape in consideration. Recently, however, there has been considerable interest in notions of scaling and self similarity, including attempts to characterize landscapes and other morphological features by fractal dimensions. Fractals generally arise due to a lack of fundamental scales, something that is not the case if we are still to believe D_d is an important parameter.

We present a theory for the explanation of drainage density in terms of the physical processes present in the landscape. The basic length scale is recognized as a point of transition between the stable smoothing effects of diffusive processes at small scale and the unstable effects of concentrative processes, such as overland or open channel flow, at large scale.

Correspondence to: D.G. Tarboton, Utah Water Research laboratory, Utah State University, Logan, UT, USA.

Since different scaling regimes exist below and above this transition point, the theory offers insight about this scaling.

Literature review

There has been considerable work done on quantifying the structure and scale of river networks among which are the notions of stream order (Horton, 1932, 1945; Strahler, 1952, 1957) and magnitude (Shreve, 1966). These are topological, dimensionless, measures of size and need to be related to the physical size (area) of the basin. This relationship is through the drainage density (eqn. 1) which is a measure of the degree to which a basin is dissected by channels. Horton (1945) suggested that the average length of overland flow or hillslope length could be approximated by $\frac{1}{2}D_d$. Horton also related drainage density to the length of first-order streams, basin area, and stream length and bifurcation ratios. Smith (1950) measured the fundamental scale of topography in terms of a texture ratio — the number of contour crenulations divided by contour length. Smith essentially showed that texture was correlated with D_d so the notion of a well or poorly drained basin corresponds to the notion of fine or coarse texture. Melton (1958) showed that stream frequency (the number of streams per unit area) was strongly correlated with drainage density. Others (Shreve, 1967; Smart, 1978) have related mean link length and link frequency (the number of links per unit area) to drainage density. Thus, there are many roughly equivalent measures of a basic length scale associated with the dissection of the landscape by the river network. The determination of this length scale is generally dependent on the resolution of the map used. Historically, workers have called for the highest resolution maps and/or field work to measure these quantities.

Mark (1983) discusses the differences between drainage networks obtained from maps and field surveys and the merits of various procedures such as use of contour crenulations to "extend" the network. He concludes that first-order basins defined from contour crenulations on 1:24,000 maps do exist as topographic features in the field. However, the form has often been simplified by cartographic generalization. Most first-order basins defined on the map contain more than one fluvial channel in the field. Accordingly, the exterior links drawn by contour crenulations do not represent unbranched channels. However, a question arises in the context of scaling and fractals (Mandelbrot, 1983) as to whether this notion of scale is well founded or whether the river networks dissect the landscape infinitely, requiring characterization as a scaling phenomena. This idea was recognized early by Davis (1899, p. 495) who wrote:

> "Although the river and hillside waste do not resemble each other at first sight, they are only the extreme members of a continuous series and when this generalization is appreciated one may fairly extend the 'river' all over its basin and up to its very divide. Ordinarily treated the river is like the veins of a leaf; broadly viewed it is the entire leaf".

The scaling in channel networks has traditionally been described in terms of Horton's ratios (Horton, 1932, 1945; Strahler, 1952, 1957). The ratio of number of streams, length of streams, are of streams and slope of streams between successive orders is approximately constant. A semi-log plot of the number, length, area and slope of streams against order is roughly a straight line. The ratio or "Horton number" is obtained from the slope of the straight line fit to such plots, the procedure being called a "Horton analysis." Mathematically the ratios are:

$$R_b = N_{w-1}/N_w, \quad R_l = L_w/L_{w-1}, \quad R_a = A_w/A_{w-1}, \quad R_s = S_{w-1}/S_w \tag{2}$$

where N_w is the number of streams of order w, L_w is the mean length of streams of order w, A_w is the mean area contributing to streams of order w, and S_w is the mean slope of streams of order w. R_b, R_l, R_a, and R_s are bifurcation, length, area and slope ratios, respectively. Since the ratios are approximately constant within a drainage network, the above geometric descriptors are called "Horton's laws." The area law above was not explicitly stated by Horton, and is due to Schumm (1956). Leopold and Miller (1956) extended and explained Horton's ideas by showing that the log of many hydraulic variables are approximate linear functions of basin order. This behavior is due to the fact that most quantities depend strongly on the size of the drainage basin, measured in terms of basin area. The relationship is often obtained from a straight line fit on a log–log plot, i.e., a power law. The fact that the area law relates order to the log of the size measure, area, leads to Horton's laws. It has been shown previously (La Barbera and Rosso, 1989; Tarboton et al., 1988, 1990; Tarboton, 1989) how the planform scaling described by the bifurcation and length ratios can be used to determine the fractal dimension of channel networks..

It is widely recognized that elevation, related to potential energy, is an important part of the network and there is a need to understand the structure and scaling of river networks with the third dimension elevation included. Qualitatively, streams are steep near their sources and flatter downstream. Horton's slope law quantifies this:

$$S_w = (R_s S_1) R_s^{-w} = R_s S_1 \exp(-w \ln R_s) \quad (3)$$

This is an exponential decrease of slope with order. Flint (1974), building on the power law relationships of Wolman (1955), Leopold and Maddock (1953), Leopold and Miller (1956), and Leopold et al. (1964), finds slope, S, empirically related to the contributing area, A, by:

$$S = CA^{-\theta} \quad (4)$$

where C is a constant and the exponent θ ranges from 0.37 to 0.83 with a mean of 0.6. Substituting in Horton's area and slope laws, Flint (1974) obtains:

$$\theta = \ln R_s / \ln R_a \quad (5)$$

This again shows the connection between power law scaling with area and exponential scaling with order (Horton's slope law).

This sort of power law scaling is the focus of considerable recent research in the context of self similarity, fractals, multifractals and self-organized critically (Bak et al., 1987; Gupta and Waymire, 1989; Tarboton et al., 1989; Rodriguez-Iturbe et al., 1992a, b).

Mandelbrot (1977) found that many natural lines, such as coastlines, contours, political boundaries (sometimes consisting of rivers or coastlines), etc., seemed to have fractal dimensions near 1.2–1.3. He also noted that simulations of fractional brownian surfaces with $D \simeq 2.3$ looked remarkably like the real landscape. He took this as evidence that landscapes were fractal characterized by $D \simeq 2.3$. Following Mandelbrot the fractal dimension of topography has been investigated by many (Mark and Aronson, 1984; Ahnert, 1984; Culling, 1986; Culling and Datko, 1987; Matshushita and Ouchi, 1989; Goodchild and Mark, 1987; Gilbert, 1989; Huang and Turcotte, 1989, 1990; Goff, 1990), using diverse techniques (variograms, Richardsons method, Fourier transforms, etc.). In many cases there is nonlinearity which has led to the suggestion that fractal models are questionable (Gilbert, 1989) or limited to a range of scales (Goodchild and Mark, 1987). There are sometimes distinct domains with different fractal dimensions (Mark and Aronson, 1984; Culling and Datko, 1987, fig. 6; Matshushita and Ouchi, 1989) which suggests a possible interpretation in terms of different processes operating or dominating at different scales. However, the link between fractal dimensions and processes is not straightforward. Newman and Turcotte (1990) suggest a model for the scaling of landscapes based on a cascade in the fourier wavelength space, somewhat analogous to historical models of turbulence. The physical basis for such a model is not yet clear.

When considering channel network and landscape geometry, it is important to have an appreciation of the processes that have and are continuing to sculpt the landscape. Ultimately, the objective is to understand the relationship between geomorphological form and processes and to make quantitative statements about the processes from detailed analysis of the form. Hydrologists are particularly interested in runoff processes and movement of water; their interest in geomorphology is based upon a need to address the problem of prediction from ungauged basins. By embracing geomorphology, hydrologists are attempting to deduce processes from land form and channel network morphology.

It has been suggested that the evolution of channel networks and hillslopes may be thought of as an "open dissipative system" (Leopold and Langbein, 1962; Scheidegger, 1970; Thornes, 1983; Hugget, 1988). Carson and Kirkby (1972) provide a good review of the early work on evolution of hillslopes, relating form to processes. Kirkby (1971) and Smith and Bretherton (1972) write equations for the evolution of landscapes based on conservation of mass with the assumption that sediment is transported in a downslope direction and surface runoff is generated uniformly. Based on these, Smith and Bretherton (1972) use a linear stability analysis to show that landscape evolution is unstable in the sense that small perturbations grow when:

$$F - q\,\partial F/\partial q < 0 \tag{6}$$

where F is the sediment flux and q surface flow. Smith and Bretherton (1972) also show that a one-dimensional equilibrium or constant form solution is concave when eq. (6) is satisfied and convex when it is not. Therefore, there is an equivalence between concavity of a one-dimensional profile and instability in the two-dimensional landscape. Instability, as characterized above, would lead to rilling and channel growth, otherwise, smooth hillslopes would prevail. Loewenherz (1991) re-evaluates the Smith and Bretherton stability criterion, demonstrating its rigor and validity. She also suggests a short wavelength cutoff to represent the lower scale limit of validity of the continuum equations for surface deformation. This results in a slight shift in the stability threshold.

The Smith and Bretherton (1972) formulation which assumes the constitutive function $F(S, q)$, parameterizes the sediment transport in terms of slope and flow which is related to upslope area. This is most appropriate in a transport limited situation where the sediment load is in equilibrium with slope and flow. F is actually the sediment transport capacity of the surface flow. Where the removal of sediment is limited by weathering and hillslope development depends on variations in weathering rate another approach is needed. Carson and Kirkby (1972) discuss many of the factors which affect weathering rate. Luke (1972, 1974) generalizes the Smith and Bretherton formulation by instead treating sediment flux, or load, as a variable and including an additional equation to represent the erosion or depositon of sediment. Luke (1974) shows how under conditions of instability, troughs develop into shock discontinuities, interpreted as channels. However, no general condition for instability of his set of equations, analogous to eq. (6) above is known.

The stability behavior of the conservation equations given by Smith and Bretherton (1972) is clearly dependent on the form of the sediment transport flux function $F(S, q)$. If flux F is proportional to S we get the widely studied linear diffusion equation which Culling (1960, 1963, 1965) analyzed in the context of slope development. If F is dependent on S nonlinearly, there is nonlinear diffusion. In fact, if the main contribution to variation of F is S, then the process will be predominantly diffusive; whereas if the main contribution to variation of F is q, the flow volume related to the carrying capacity of the concentrated flow, the process is advective. In diffusive processes F is predominantly dependent on slope S and the term $\partial F/\partial q$ in eq. (6) is small, which leads to eq. (6) being positive and, consequently stability of the landscape evolution under this process. However, if the process has an advective component, there is always some flow large enough to make the term $q(\partial F/\partial q)$ dominate eq. (6) and result in instability. The effects of convergence and concentration of flow eventually dominate. Kirkby (1987) explains this instability:

> "Where flows of water and sediment converge, the combined water flow drives a sediment transport greater than the sum of the sediment inflows, so that there is an excess transporting capacity which is able to enlarge the hollow."

Thus advective processes result in channel initiation and development.

A common form of F used to parameterize many processes is (Kirkby, 1971; Table 1):

$$F = \beta a^m S^n \quad (7)$$

where β, m and n are coefficients and "a" is upslope area, i.e., the area drained per unit contour width. Kirkby (1971) takes discharge q as proportional to "a" with an exponent of 0.75 to 1. Table 1, excerpted from Kirkby (1971), gives typical m and n for various processes. Mathier et al. (1989) and Julien and Simons (1985) also discuss overland flow sediment transport parameterizations of this form. Kirkby (1971) studied the solution to the continuity equations in one dimension and identified characteristics profiles under different

TABLE 1

Typical values of exponents m and n in the empirical relationship $F = \beta a^m S^n$

Process	m	n
Soil creep	0	1.0
Rain splash	0	1–2
Soil wash	1.3–1.7	1.3–2
Rivers	2–3	3

From Kirkby (1971).

parameter values. These show convex profiles for soil creep progressing to concave profiles for rivers.

Kirkby (1980) focuses on the instability criterion of Smith and Bretherton (1972) and notions of dominant domains and transition thresholds. He suggests that when the landscape is unstable, small hollows will grow into valley heads. He shows that where a combination of creep and wash sediment transport processes are present, there is a critical area a_c where instability [according to eq. (6)] occurs. He suggests that for a landscape with fully developed drainage, the area drained per unit length of channel bank must be less than a_c, which results in a lower bound on drainage density:

$$D_d > 1/2a_c \quad (8)$$

Kirkby suggests that for efficient networks, this approaches equality and provides a way to estimate drainage density. The notion of a critical area or critical hillslope length is also recognized by Dunne (1980) who points out that sheet flow can still occur in the stable domain and rilling and channelization occur some way beyond the point (down a hillslope) where sheet flow occurs. Dunne suggests that this may be because of the diffusive, leveling influence of rainsplash. The instability occurs where the unstable sheet flow dominates transport. Thornes (1983) also emphasizes the importance of defining domains where different processes dominate and what the transitions between domains are. An additive form to the sediment transport function can be used to represent the sum of different processes. Smith and Bretherton (1972) and Band (1985) used the form:

$$F = KS + \alpha q^m S^n \quad (9)$$

where K and α are constants. This is a sum of creep and wash processes. Band (1985) notes that the relative magnitude of diffusive and surface wash rates are crucial to the resulting slope shapes. Kirkby (1987) uses this equation with runoff generated according to the topographic model of Beven and Kirkby (1979).

Kirkby (1987) also gives a formulation for a combination of landslide and creep or splash sediment removal, resulting in concave and, presumably, unstable profiles beyond a critical distance downslope. Landsliding is formulated as an erosion limited process. This raises a question on the validity of the interpretation of concavity as implying instability because the sediment flux is not in the form $F(S, q)$ required to apply the Smith and Bretherton (1972) result. Sediment flux could be written in Luke's (1972, 1974) more general formulation but to our knowledge stability of the resulting more general system has not been investigated. Kirkby (1987) also mentions a difficulty with this formulation is that it does not require hollows to be occupied by streams, something that is usually the case in field observations.

Andrews and Bucknam (1987), in the context of scarp degradation, have suggested:

$$F=K_0\frac{S}{1-(S/f)^2} \tag{10}$$

where f is a friction or sliding slope and K_0 is a constant. This form is derived from consideration of the distance a particle travels before sliding to a stop after it has been given an initial velocity (perhaps from a raindrop or animal hoof). This is a nonlinear diffusion term that could be used instead of KS in equation (9).

Ahnert (1987) describes a slope denudation model developed over the past three decades. This is based on the continuity equation with the added feature that sediment removal is a function of regolith (i.e., soil cover) thickness. This allows the modeling of weathering-limited and transport-limited situations. The effect of weathering-limited denudation is only explored in a limited way. Ahnert (1987) also notes that the surface wash form of sediment transport results in concave slope profiles and diffusive sediment transport results in convex profiles. He suggests that the concave (wash dominated) part of a profile represents the longitudinal profile of streams and the convex part (shaped by diffusive mass movements) represents the valley side slopes.

Recently Willgoose (1989 and Willgoose et al., 1991a, b) developed a catchment and channel network evolution model. The model is based on sediment transport continuity but postulates an explicit difference between sediment transport on a hillslope and in a channel. This is implemented numerically by using a sediment transport function F, dependent on an index of channelization Y. Y is an indicator variable taking the value 1 in channels and 0 on hillslopes. Willgoose uses the following parameterization of transport:

$$F(S,q,Y)=\begin{cases}\beta O_t q^m S^n & Y=0\\ \beta q^m S^n & Y=1\end{cases} \tag{11}$$

where again β, m and n are coefficients. The factor O_t gives the relative differences between sediment transport on hillslopes and in channels. Willgoose suggests that channelization occurs (the switch from $Y=0$ to $Y=1$) when a channel initiation function exceeds a certain threshold. He uses sediment transport arguments to suggest an initiation function of the form:

$$CI(S,q)=\beta' q^{m'} S^{n'} \tag{12}$$

where β', m' and n' are coefficients, in general different from those in eq. (11). Willgoose (1989 and Willgoose et al., 1991a,b) show that simulations based on this model result in realistic looking channel networks. He compared Horton (1945) and Tokunaga (1978) scaling ratios as well as hypsometric properties to demonstrate how well the model reproduces properties of natural channel networks. Willgoose et al. (1990) also showed that the model reproduced the slope–area scaling at channel heads observed by Montgomery and Dietrich (1988). The notion of sediment transport rate being different on hillslopes and channels is intuitively appealing, although it has not been observed in the field or experimentally justified. The date of Priest et al. (1975) used by Willgoose (1989) suggest $O_t = 0.4$ which may not be significantly different from $O_t = 1$. The Willgoose model is distinct in that channels are imposed when the channel initiation criterion is met. This needs not in general coincide with the development of shock discontinuities described by Luke (1974) or the instability threshold of Smith and Bretherton (1972), eq. (6).

A theoretical model for landscape evolution

Here we suggest that a reformulation of the continuity equations suggested by Smith and Bretherton (1972) and Kirkby (1971) with sediment flux F, a function of slope S and upslope area "a", rather than flow q, provide minimum necessary complexity sufficient for instability, the formation of discontinuities and channel networks. An assumption of dynamic equilibrium leads to a relationship between slope and upslope area. This is a scaling function which characterizes the landscape and has a form dictated by the landform evolution processes. We show how different sediment transport processes manifest themselves in terms of this scaling. Where multiple processes are present, changes in upslope area can result in a switch in the dominating process and a corresponding switch in the stability criterion (Smith and Bretherton, 1972) indicating a threshold upslope area for the formation of channels. An advantage of the formulation in terms of a slope–area scaling function is that it can be measured using topographic or digital elevation model (DEM) data. The slope–area scaling function dictates a support area with which channel networks should be extracted from DEM's and hence gives a drainage density.

The basic equations which we regard as sufficient to model the essence of landscape evolution are:

$$\partial z/\partial t = -\nabla \cdot \underline{n} F(|\nabla z|, a) \qquad (13)$$

and:

$$\nabla \cdot \underline{n} a = 1 \qquad (14)$$

with:

$$\underline{n} = -\frac{\nabla z}{|\nabla z|} \qquad (15)$$

where z, elevation, and a, upslope area, are the dependent variables that vary over space (x, y) and time t and $\underline{n}$ is a unit vector in the downslope direction. The equations are conservation of sediment mass (eq. 13) and calculation of upslope area (eq. 14). The definitions of downslope directions (eq. 15), and the sediment transport function, $F(|\nabla z|, a) = F(S, a)$, are constitutive relationships. Figure 1 illustrates the definition of area per unit contour width or upslope area "a". The only difference between these equations and those of Smith and Bretherton (1972) is that the independent variable q in $F(S, q)$ has been replaced by "a". This can be done by simple substitution if one assumes there exists a function $q(a)$:

$$F(S, q(a)) = F(S, a) \qquad (16)$$

Kirkby (1971) used a power function for $q(a)$:

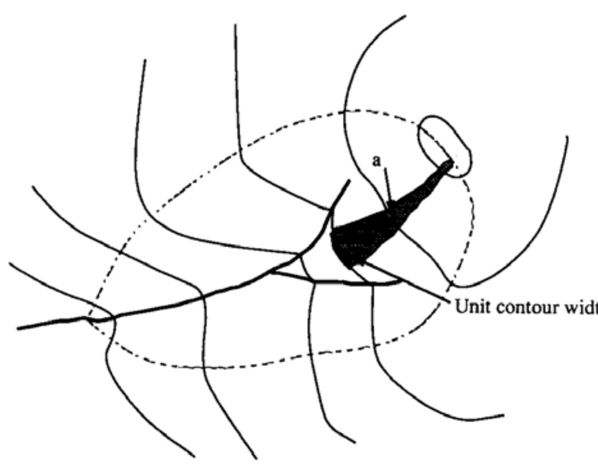

Fig. 1. Sketch defining partial catchment area per unit contour with "a".

however different, perhaps more complicated functions may be required to model the partial response of hillslopes with variable contributing areas. The $F(S, a)$ construct also allows more generality since surface flow could be functionally dependent on slope, i.e. $q(S, a)$, to perhaps model slope dependent runoff generation mechanisms. All parameterizations of different sediment transport processes are incorporated in the function $F(S, a)$ which dictates the landforms and drainage density that results. In general, we would expect the sediment flux to also be a function of many other variables perhaps including vegetation, erosion resistance and whether or not the flow is channelized. Then one could write $F(S, a, R, V, Y, ...)$. However, here we assume that these effects are subservient to the effects of slope and upslope area which are the focus of our analysis. The other variables may be responsible for some of the apparently random scatter in the results.

An assumption of dynamic equilibrium, or constant average degradation rate, seems reasonable for a sufficiently localized region and has been widely used (Davis, 1899; Hirano, 1975). Within the present framework this is equivalent to assuming that the sediment transport past every point is proportional to upslope area (Willgoose et al., 1991c):

$$F(S, a) = U\,a \tag{17}$$

For U given and the functional form $F(\cdot)$ known, this is an implicit relationship between slope S and area a. In this expression, U does not have to be restricted to uplift but can be thought of as average degradation rate. With $F(\cdot)$ given by $\beta a^m S^n$, as in eq. (7), this relationship can be explicitly solved:

$$S = \left(\frac{U}{\beta}\right)^{1/n} a^{-(m-1)/n} \tag{18}$$

This is a power law scaling of slope with area and has been used by Willgoose et al. (1991c) to explain the empirical observations described by eq. (4). For a more general $F(\cdot)$, recognizing that eq. (17) defines a function $S(a)$ implicitly, we can write:

$$F(S(a), a) = U\,a \tag{19}$$

which upon differentiation and multiplication by "a" gives:

$$a\frac{\partial F}{\partial S}\frac{\mathrm{d}S}{\mathrm{d}a} + a\frac{\partial F}{\partial a} = U\,a = F \tag{20}$$

This can be rewritten:

$$a\frac{\partial F}{\partial S}\frac{\mathrm{d}S}{\mathrm{d}a} = F - a\frac{\partial F}{\partial a} \tag{21}$$

Here the right-hand side is equivalent to the stability criterion of Smith and Bretherton (1972). For $F - a(\partial F/\partial a) < 0$, small perturbations grow into rills and ultimately channels. For $F - a(\partial F/\partial a) > 0$, small perturbations do not grow so the landform remains smooth, i.e., hillslopes. Now, on the left-hand side of eq. (21), "a" is always positive and $\partial F/\partial S$ we expect to be positive so stability depends on the sign of the gradient $\mathrm{d}S/\mathrm{d}a$. $\mathrm{d}S/\mathrm{d}a < 0$ results when there is instability and channelization, otherwise $\mathrm{d}S/\mathrm{d}a > 0$. This suggests that a break or change in gradient of the $S(a)$ function is a dividing scale, separating the distinctly different regimes of channels and hillslopes. (Note: The term *gradient* is used for the slope of a graph or function and the term *slope* for the slope of the ground.) Equation (18) considered only one sediment transport function under dynamic equilibrium. In principle, many sediment transport processes may operate at the same scale (Tarboton, 1989; Willgoose, 1989; Willgoose et al., 1991c). Consider the presence of two mechanisms and assume that the total sediment flux is the sum of contributions from each mechanism:

$$F(S, a) = F_1(S, a) + F_2(S, a) \tag{22}$$

Equating the above to Ua as in eq. (19) and solving for $S(a)$, the slope–area profile under dynamic equilibrium for combined sediment transport is obtained. This is given in Fig. 2 for three different plausible mass wasting type

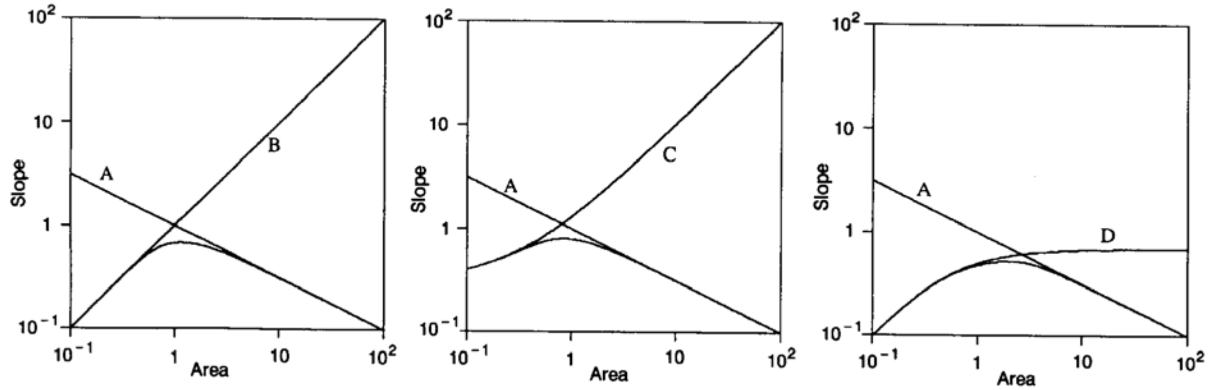

Fig. 2. Equilibrium slope profiles for combinations of transport functions. The following four sediment transport functions in dimensionless units are assumed:
A: $F=a^{2.5}S^3$ for rivers (Leopold and Maddock, 1953; Kirkby, 1971); B: $F=S$ for soil creep (Kirkby, 1971); C: $F=S-0.3$ for landsliding (Kirkby, 1971); D: $F=S/[1-(S/0.7)^2]$ for sliding (Andrews and Bucknam, 1987).
Each figure shows the solutions to: $F(S,a)=a$ for each of the two mechanisms selected, and the solution to: $F_1(S,a)+F_2(S,a)=a$. These represent the slope–area scaling under dynamic equilibrium for each single mechanism, and the resultant slope area scaling if the two mechanisms exist together.

sediment transport functions combined with a wash sediment transport function. This figure shows that in all three cases a slope–area profile that changes from a positive, or near zero gradient, to a negative gradient is obtained. This change over is accompanied by a switch in the process that dominates the sediment transport. At small contributing areas, or small scale, the sediment transport is dominated by the mass wasting process; whereas, for large contributing areas or scale, wash erosion dominates and due to the instability the landscape will become channelized.

Thus, the break in slope–area scaling represents the transition point between hillslope and channelized regimes. This analysis points out the importance of slope–area plots in the analysis of landforms. These plots are the signature of processes on the landscape, provide information about the fundamental scales, and suggest how the scaling in between these scales can be interpreted in terms of sediment transport processes.

Examples using digital elevation data

An advantage of this perspective on landscapes is that slope and upslope area can be measured using digital elevation model (DEM) data. Tarboton (1989) and Tarboton et al. (1991) describe techniques to detect the break in slope–area scaling and estimate the corresponding drainage density for 21 U.S. Geological Survey grid based DEM datasets comprised of several adjacent DEM quadrangles in most cases. These were selected to represent a broad range of landscape types, but were constrained by the availability and cost of DEM data for complete river basins. To understand the analysis and processing of DEM's it is necessary to define the procedures and terminology used, which follows the work of O'Callaghan and Mark (1984), Mark (1988) and Jenson and Domingue (1988).

Elevations are stored in an *elevation matrix* arranged in a grid with each entry giving the elevation of a point. The location within the matrix implies the spatial location of the point, so only elevation values need to be stored (as opposed to the alternative DEM data storage structures, triangular irregular networks that have to store x and y location and elevation data for each point and contour-based structures that store strings of x and y locations

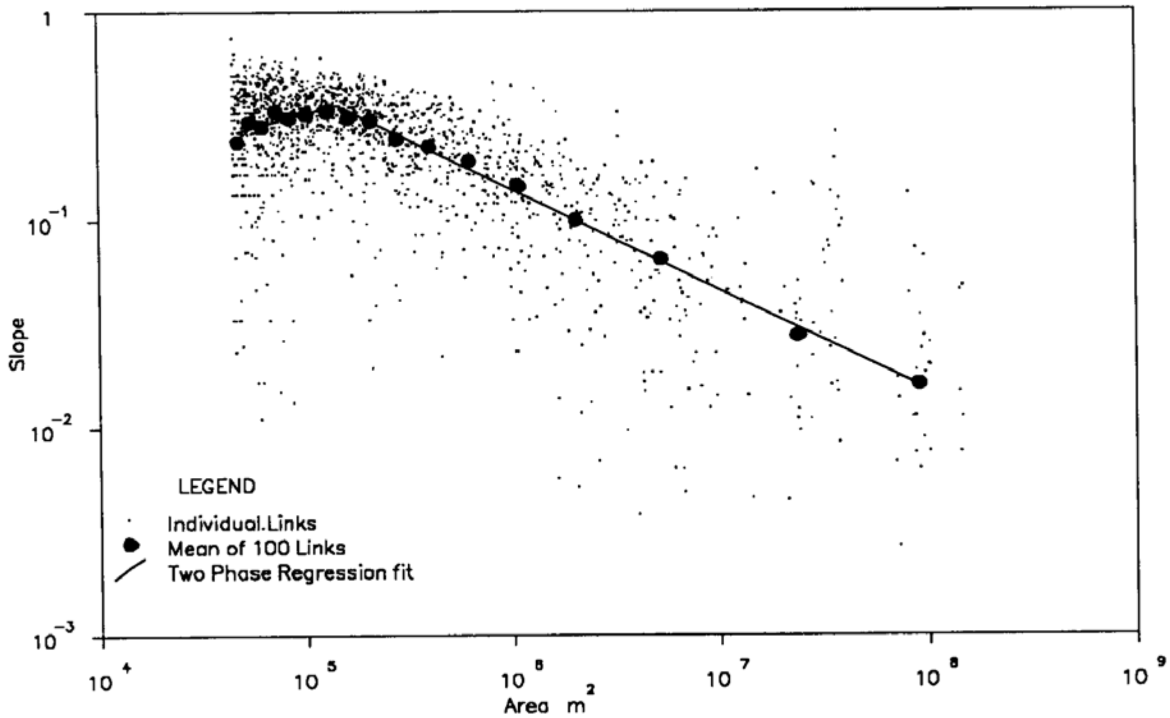

Fig. 3. Big Creek basin, Idaho: link slopes with support area of 50 pixels used to extract network.

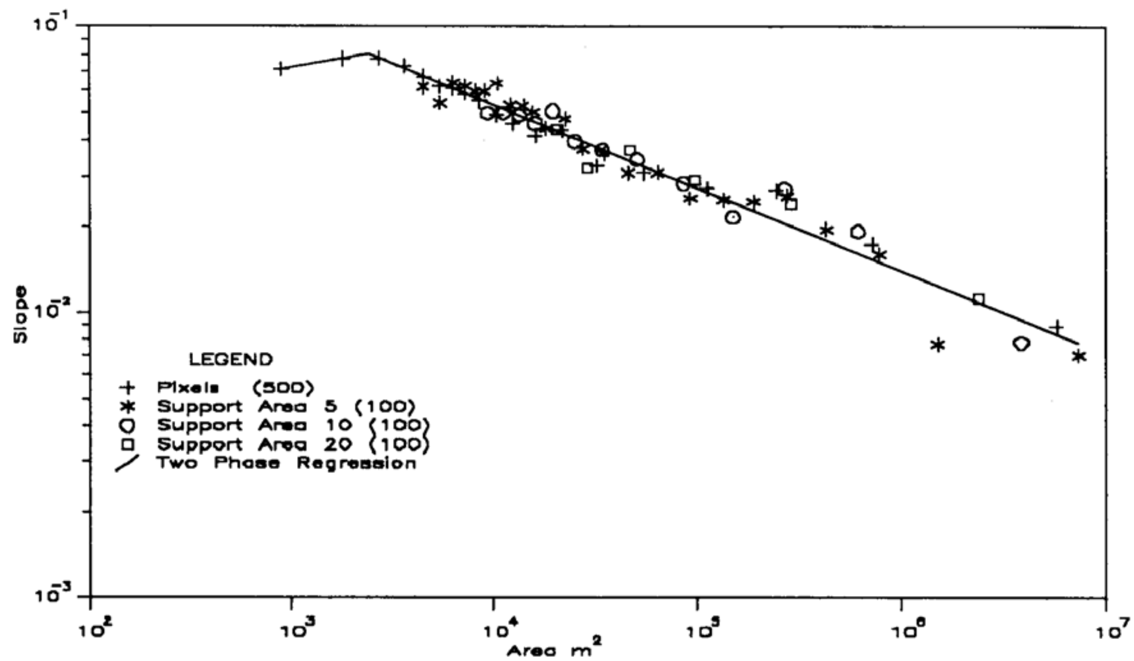

Fig. 4. W15 subbasin in Walnut Gulch Arizona: slope versus area and two-phase regression plot. The numbers in parentheses in the legend indicate the number of individual slope data values averaged together to smooth the data. (Data from U.S.G.S. 30 m resolution 7.5 minute DEM Quadrangles, Hay Mountain and Tombstone, Arizona)

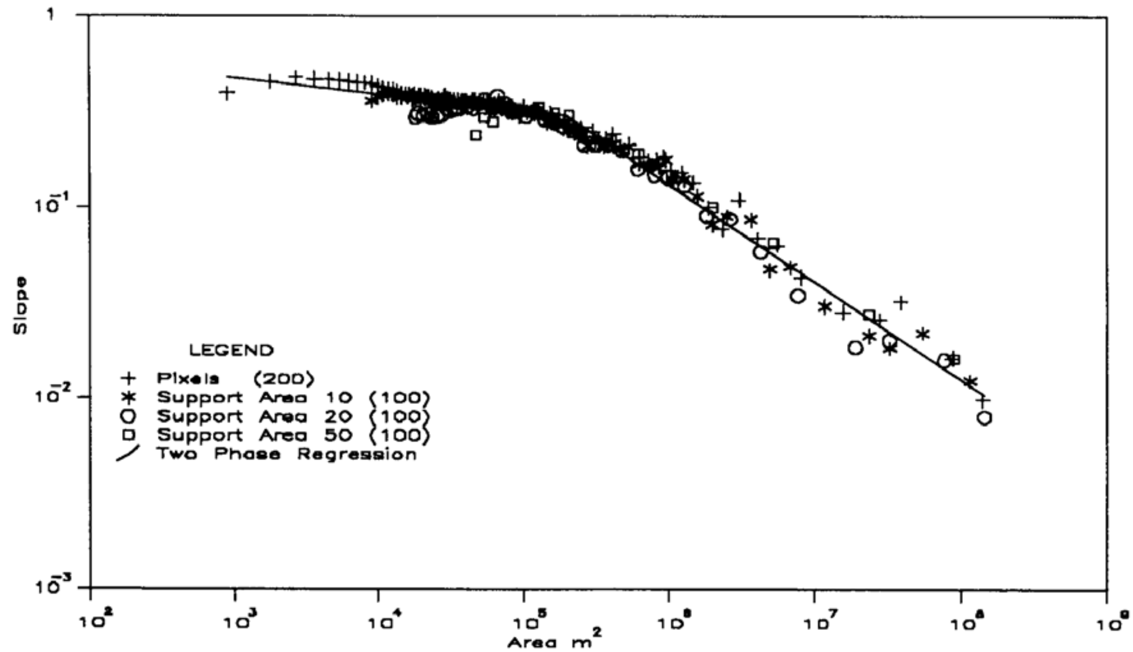

Fig. 5. Big Creek basin, Idaho: slope versus area and two-phase regression plot. The numbers in parentheses in the legend indicate the number of individual slope data values averaged together to smooth the data. (Data from U.S.G.S. 30 m resolution 7.5 minute DEM Quadrangles, Calder NE, NW, SE and SW, Idaho)

ber of points used in the moving average indicated in parentheses in the legend. Some problems occur with this procedure. First, the break in gradient is often not distinct; in some cases (e.g. Fig. 6) it appears to be dependent on support area threshold and, hence, slope averaging length. Also, the break in gradient of the overall regression is often not to a positive gradient at small upslope area, although if a single averaging length were used it would be.

These problems may be due to errors in the DEM data, so to understand the effect of errors in our procedures, we constructed a data set with predetermined slope–area functions and applied random additions of noise. A one-dimensional hillslope profile was specified as:

$$z=250-x^2/9000 \tag{23}$$

where z is elevation and x distance from the divide. For a one-dimensional profile $x=a$, and the slope scaling is:

$$S=\left|\frac{\mathrm{d}z}{\mathrm{d}x}\right|=a/4500 \tag{24}$$

Equation (23) was used to compute elevations on a 51×102 grid with spacing of 30 m schematically shown in Fig. 7. A random measurement error simulated from a zero mean Gaussian distribution with variance up to four was added to each elevation and then rounded to the nearest meter for storage as an integer value as with real DEM data. Slope–area profiles were then computed using the same procedures as for digital elevation models. As illustrated in Fig. 8, the results show, even for the relatively small errors simulated, a large effect on the slope–area relationship. The theoretical line is obtained from eq. (24) using a pixel size of 30 m. Area per unit contour width "a" is calculated as contributing area A divided by 30 m. Thus the theoretical line in Fig. 8 is given by:

$$S=A/30/4500=A/135{,}000 \tag{25}$$

The good data is with no added measurement noise and has scatter due to the rounding of

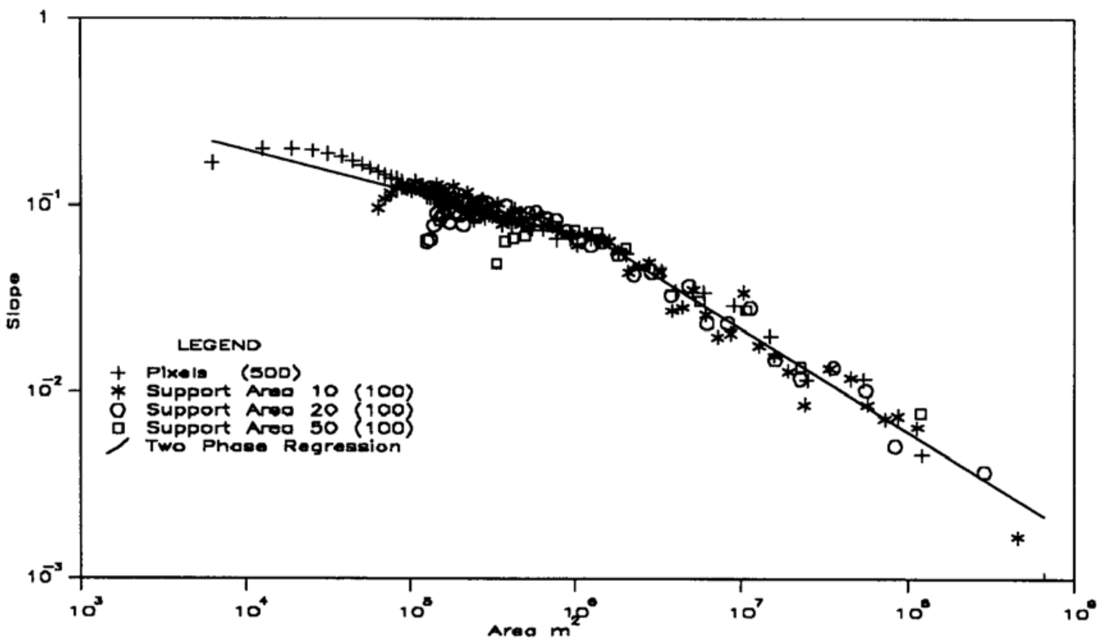

Fig. 6. East Delaware basin: slope versus area and two-phase regression plot. The numbers in parentheses in the legend indicate the number of individual slope data values averaged together to smooth the data. (Date from D.M.A. 3 arc second resolution (60×90 m) 1°×1° DEM Quadrangle, Binghampton, New York)

elevations to the nearest meter. In the noisy data, the errors introduce a negative correlation between slope and area which tends to reduce the positive slope of the slope–area curves. This may cause the negative slopes in the slope–area data (Figs. 4 to 6), to the left of the break in scaling, at scales or areas we interpret as hillslopes.

To understand this effect, consider an error in the elevation of a single pixel. If the error reduces the apparent elevation of a pixel, the apparent slope is reduced. This is because the

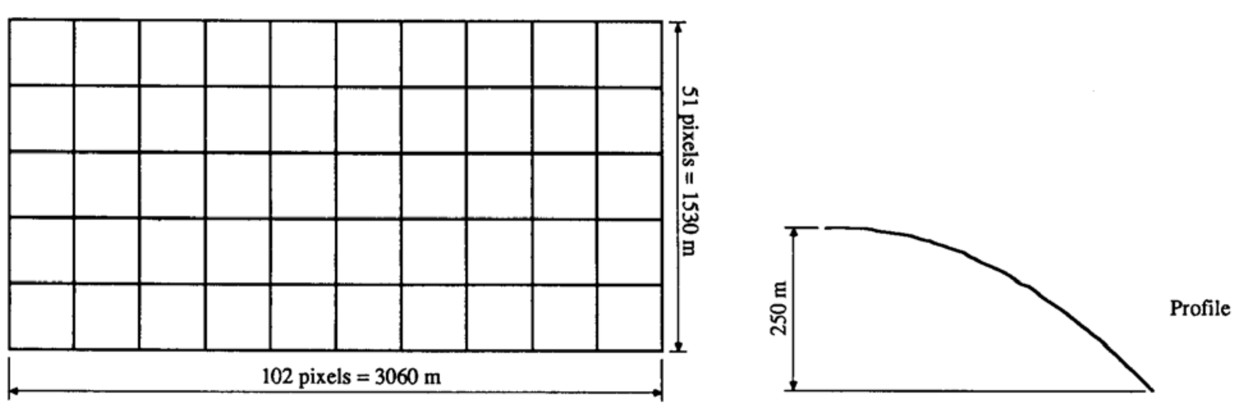

Fig. 7. Schematic diagram of simulated slope profile.

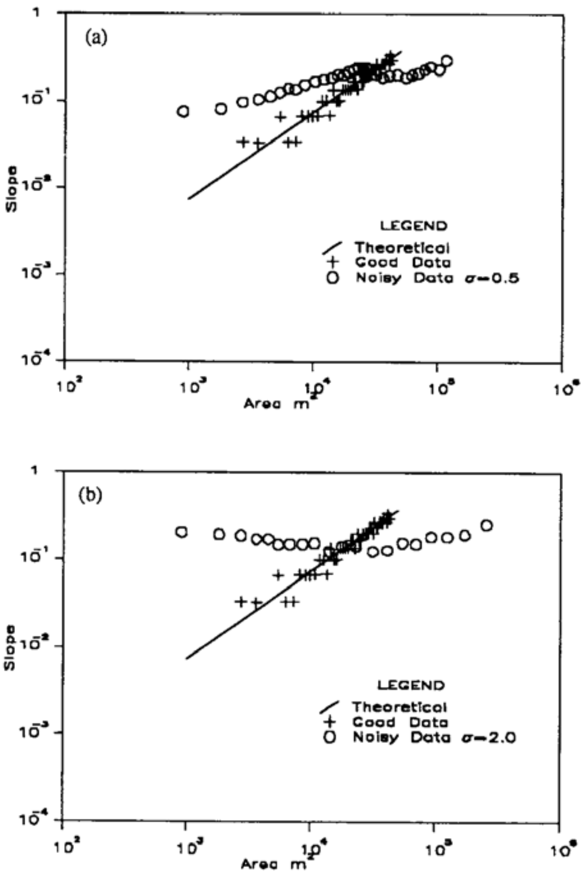

Fig. 8. Slope–area plots for simulated hillslope profile. (a) Added noise standard deviation = 0.5 m; (b) Added noise standard deviation = 2.0 m. Note: Data with slope = 0 omitted from graph.

slope of a pixel is measured as the difference in elevation between the pixel in consideration and its downslope neighbor, divided by the distance between pixels. Also, adjacent pixels are more likely to drain towards the pixel in consideration due to its reduced elevation, thus increasing the apparent area that it drains. Similarly, an error that increases the apparent elevation, increases slope, and reduces area, so the net effect is that errors result in a negative correlation between slope and area or negative gradient in slope–area plots.

In some of the results, Fig. 6 in particular, there is a strong dependence of the break in scaling on support area or averaging length. We have not yet fully resolved the causes for this but think that the explanation lies in issues of resolution and bias introduced by the sampling procedures. In comparing, for example, the asterisks and circles on Fig. 6 we are comparing slopes calculated from links with support areas of ten and twenty pixels. When a support area of twenty pixels is used the links that have support area just larger than twenty will be first order and may be very short, leading to uncertainty in the estimated slope. When the support area is reduced, to say ten pixels, these short links are extended upwards to include more pixels that have smaller area and hence steeper slopes. This would have the effect of eliminating the "hook effect" near the threshold of twenty. However it would still be a problem near the new support area threshold of ten. It is worth noting that this problem is most severe for the East Delaware river basin which is from a low-resolution 3 arc second (60×90 m) DEM dataset and is not as noticeable in the higher-resolution datasets. Further work with high-resolution data is probably warranted to fully resolve these issues.

The break in slope–area scaling gives a procedure for obtaining channel support area and, hence, drainage density that is physically justifiable in terms of the hillslope transport processes. Tarboton (1989) and Tarboton et al. (1991) showed that, despite the errors, drainage density obtained in this manner compares fairly well to drainage density obtained using other techniques.

Conclusions

The fact that there are fundamental or basic scales where the slope–area scaling breaks suggests different processes above and below the break. The difference between stable diffusive processes and unstable channel-forming processes, under the Smith and Bretherton (1972) stability criterion has been shown to be equivalent to a change in the sign of the slope-area scaling function gradient. Thus, the break can be interpreted as the scale at which stability changes and as such can be used to determine drainage density. Where there are multiple sediment transport processes present, the break point is the point where domination by stable diffusive processes yields to domination by unstable channel-forming processes.

Much of fluvial geomorphology has focused on the slope–area scaling relation $S \sim A^{-\theta}$ which may be explained in terms of dynamic equilibrium and wash processes (eq. 18). This cannot hold down to the limit of zero area as it implies infinite slope. Diffusive slope-dependent transport mechanisms will always dominate as area gets small and slope gets large, resulting in a limit to this scaling. This transition may, however, be at a scale too small to resolve, given the data resolution. Thus, it is only useful to characterize the landscape as scaling above or below this limit or basic scale.

On either side of the basic scale the differential equations governing landform evolution are in different stability regimes. Below the basic scale they are stable, while above it they are unstable. Therefore, we should expect fundamentally different scaling behavior in the two scaling regimes. Perhaps this explains the change of fractal dimension with scale that has been reported by others (Mark and Aronson, 1984; Culling and Datko, 1987).

Here we have shown that considerable understanding of landscape structure can be gleaned from the parameterization of sediment transport in terms of the topographic variables, slope, and upslope area. Analysis in terms of the slope–area scaling function can be done theoretically for specified processes and practically in terms of measured slope and area. A break in this scaling function indicates different regimes of stability and instability and can be used to determine drainage density. The connection between drainage density and the land-forming processes is one more step towards understanding the link between processes and morphology.

This work has taken the perspective that the stability threshold defines drainage density and the location of channel heads within a first-order basin. The modeling of Willgoose et al. (1991a–c), however, takes an alternative perspective. Channel heads and consequently drainage density are controlled by an activator function not related to the stability threshold. Slope–area scaling with a negative exponent can (and in their simulations does) still exist on hillslopes above the channel head in their perspective. Future work with perhaps higher-resolution data and field measurements of transport processes may be required to resolve this point.

This work has provided a physical basis for distinct scaling regimes in the landscape and the land-forming processes under the transport-limited formulation presented. Future work should extend the stability analysis to a more general formulation that includes erosion-limited processes. Future work should also investigate the relationship between landscape fractal dimension and area–slope scaling within the stable and unstable scaling regimes. The multiscaling of slopes with area (Tarboton et al., 1989) still lacks a physical explanation. The notion of self-organized criticality (Bak et al., 1987, 1988; Hwa and Kardar, 1989) has been used to describe the physics of fractals in other contexts. The idea is that power law scaling and fractal properties arise as the minimally stable states of dynamical systems with extended spatial degrees of freedom. These minimally stable states are called critical states by the analogy with thermodynamics and the scaling behavior of substances near the thermodynamic critical point. There are indications that the landscape at scales larger than the drainage density scale may be in such a critical state. The governing eqs. (13 and 14) are unstable, but perturbations cannot grow without bound, due to the presence of diffusive mechanisms. Self-organized criticality is also characterized by power law distributions of energy dissipation and this has been used to characterize the three-dimensional scaling and structure of river basins (Rodriguez-Iturbe et al., 1992a, b). We feel that there are deep physical principles involving self-organization and minimum energy that are at work governing the structure of river basins and will be the key to the physical understanding of landscape fractal dimensions and scaling properties in the future.

Work in this area is somewhat limited by DEM data resolution and it is important to test the procedures for identifying basic scales on more and higher resolution datasets. Higher resolution data would be useful to test how resolution dependent the results are and to resolve whether the negative gradient in slope–area plots at small area is due to data errors, as suggested here, or is really present in the landscape, in which case other explanations need to be sought. High resolution data together with accurate measurements of transport processes is required to truly test the theoretical models of river basin evolution discussed here.

Acknowledgements

Support for this work was provided by the National Science Foundation (Grant No. ECE-8513556) and through a cooperative agreement between the University of Florence and MIT. This work has benefited from discussion with Garry Willgoose. We thank Colleen Riley for her editorial help.

References

Ahnert, F., 1984. Local relief and the height limits of mountain ranges. Am. J. Sci., 284: 1035–1055.

Ahnert, F., 1987. Process-response models of denudation at different spatial scales. In F. Ahnert (Editor), Geomorphological Models, Theoretical and Empirical Aspects. Catena, Suppl. 10: 31–50.

Andrews, D.J.and Bucknam, R.C., 1987. Fitting degradation of shoreline scarp by a nonlinear diffusion model. J. Geophys. Res., 91(B12): 12,857–12,867.

Bak, P., Tang, C. and Wiesenfeld, K., 1987. Self organized criticality: an explanation of 1/f noise. Phys. Rev. Lett., 59(4): 381–384.

Bak, P., Tang, C. and Wiesenfeld, K., 1988. Self Organized criticality. Phys. Rev. A, 38(1): 364–374.

Band, L.E., 1985. Simulation of slope development and the magnitude and frequency of overland flow erosion in an abandoned hydraulic gold mine. In M.J. Woldenburg (Editor), Models in Geomorphology. Proc. 14th Annual Geomorphology Symposium, Binghampton, N.Y., pp. 191–211.

Beven, K.J. and Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. Hydrol. Sci. Bull., 24(1): 43–69.

Carson, M.A. and Kirkby, M.J., 1972. Hillslope Form and Process. Cambridge University Press, 475 pp.

Culling, W.E.H., 1960. Analytical theory of erosion. J. Geol., 68: 336–344.

Culling, W.E.H., 1963. Soil creep and the development of hillside slopes. J. Geol., 71: 127–161.

Culling, W.E.H., 1965. Theory of erosion on soil covered slopes. J. Geol., 73: 230–254.

Culling, W.E.H., 1986. On Hurst phenomenon in the landscape. Trans. Jpn. Geomorphol. Union, 7(4): 221–243.

Culling, W.E.H. and Datko, M., 1987. The fractal geometry of the soil-covered landscape. Earth Surf. Processes Landforms, 12: 369–385.

Davis, W.M., 1899. The geographical cycle. Geogr. J., 14: 481–504 (reproduced in Geographic Essays, edited by W.M. Davis, Ginn, Boston, 1909).

Dunne, T., 1980. Formation and controls of channel networks. Prog. Phys. Geogr., 4: 211–239.

Flint, J.J., 1974. Stream gradient as a function of order, magnitude and discharge. Water Resour. Res., 10(5): 969–973.

Gilbert, L.E., 1989. Are topographic data sets fractal. In C.H. Scholz and B.B. Mandelbrot (Editors), Fractals in Geophysics. Birkhauser Verlag, pp. 241–254.

Goff, J.A., 1990. Comment on "Fractal Mapping of Digitized Images: Application to the Topography of Arizona and Comparisons with Synthetic Images" by J. Huang and D.L. Turcotte. J. Geophys. Res., 95(B4): 5159.

Goodchild, M.F. and Mark, D.M., 1987. The fractal nature of geographic phenomena. Ann. Assoc. Am. Geog., 77(2): 265–278.

Gupta, V.K. and Waymire, E., 1989. Statistical self-similarity in river networks parameterized by elevation. Water Resour. Res., 25(3): 463–476.

Hinkley, D.V., 1969. Inference about the intersection in two-phase regression. Biometrika, 56(3): 495–504.

Hirano, M., 1975. Simulation of developmental process of interfluvial slopes with reference to graded form. J. Geol., 83: 113–123.

Horton, R.E., 1932. Drainage basin characteristics. Trans. Am. Geophys. Union, 13: 350–361.

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

Huang, J. and Turcotte, D.L., 1989. Fractal mapping of digitizing images: application to the topography of Arizona and comparisons with synthetic images. J. Geophys. Res., 94(B6): 7491–7495.

Huang, J. and Turcotte, D.L., 1990. Reply. J. Geophys. Res., 95(B4): 5161.

Hugget, R.J., 1988. Dissipative systems: Implications for geomorphology. Earth Surf. Processes Landforms, 13: 45–49.

Hwa, T. and Kardar, M., 1989. Dissipative transport in open systems: an investigation of "Self-organized criticality". Phys. Rev. Lett., 62(16): 1813.

Jenson, S.K. and Domingue, J.O., 1988. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogramm. Eng. Remote Sensing, 54(11): 1593–1600.

Julien, P.Y. and Simons, D.B., 1985. Sediment transport capacity of overland flow. Proc., Am. Soc. Agric. Eng., 28(3): 755–762.

Kirkby, M.J., 1971. Hillslope process–response models based on the continuity equation. In: Slopes Form and Processes. Inst. Br. Geogr., Spec. Publ., 3: 15–30.

Kirkby, M.J., 1980. The stream head as a significant geomorphic threshold. In: D.R. Coates and J.D. Vitek (Editors), Thresholds in Geomorphology. Allen and Unwin, pp. 53–73.

Kirkby, M.J., 1987. Modelling some influences of soil erosion,landslides and valley gradient on drainage density and hollow development. In F. Ahnert (Editor), Geomorphological Models, Theoretical and Empirical Aspects. Catena, Suppl., 10: 1–14.

La Barbera, P. and Rosso, R., 1989. On the fractal dimension of stream networks. Water Resour. Res., 25(4): 735–741.

Leopold, L.B. and Langbein, W.B., 1962. The concept of entropy in landscape evolution. U.S. Gecol. Surv., Prof. Pap. 500-A.

Leopold, L.B. and Maddock, T., 1953. The hydraulic geometry of stream channels and some physiographic implications. U.S. Geol. Surv., Prof. Pap., 252.

Leopold, L.B. and Miller, J.P., 1956. Ephemeral streams — Hydraulic factors and their relation to the drainage net. U.S. Geol. Surv., Prof. Pap. 282-A.

Leopold. L.B., Wolman, M.G. and Miller, J.P., 1964. Fluvial Process in Geomorphology. Freeman, San Francisco, Calif.

Loewenherz, D.S., 1991. Stability and the initiation of channelized surface drainage: a reassessment of the short wavelength limit. J. Geophys. Res., 96(B5): 8453–8464.

Luke, J.C., 1972. Mathematical models for landform evolution. J. Geophys. Res., 77(14): 2460–2464.

Luke, J.C., 1974. Special solutions for nonlinear erosion problems. J. of Geophys. Res., 79(26): 4035–4040.

Mandelbrot, B.B., 1977. Fractals, Form, Chance and Dimension, Freeman, San Francisco, CA.

Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. Freeman, New York, N.Y., 468 pp.

Mark, D.M., 1983. Relations between field-surveyed channel networks and map based geomorphic measures, Inez, Kentucky. Ann. Assoc. Am. Geogr., 73(3): 358–372.

Mark, D.M., 1988. Network models in geomorphology. In M.G. Anderson (Editor), Modelling in Geomorphological Systems. John Wiley, New York, N.Y., pp. 73–97.

Mark, D.M. and Aronson, P.B., 1984. Scale-dependent fractal dimensions of topographic surfaces: An empirical investigation with applications in geomorphology and computer mapping. Math. Geol., 16(7): 671–683.

Mathier, L., Roy, A.G. and Pare, J.P., 1989. The effect of slope gradient and length on the parameters of a sediment transport equation for sheetwash. Catena, 16: 545–558.

Matshushita, M. and Ouchi, S., 1989. On the self-affinity of various curves. Physica, D, 38: 241–251.

Melton, M.A., 1958. Geometric properties of mature drainage systems and their representation in an E4 phase space. J. Geol., 66: 35–66.

Montgomery, D.R. and Dietrich, W.E., 1988. Where do channels begin. Nature, 336: 232–234.

Newman, W.I. and Turcotte, D.L., 1990. Cascade Model for Fluvial Geomorphology. Geophys. J. Int., 100(3): 433–439.

O'Callaghan, J.F. and Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Comp. Vision Graphics Image Processing, 28: 328–344.

Priest, R.F. Bradford, J.M. and Spomer, R.G., 1975. Mechanisms of erosion and sediment movement from gullies. In: Proc. Sediment Yield Workshop, USDA, Sedimentation Laboratory, Oxford, Miss., November 1972, pp. 162–176.

Rodriguez-Iturbe, I. Ijjasz-Vasquez, E.J., Bras, R.L. and Tarboton, D.G., 1992a. Power-law distributions of mass and energy in river basins. Water Resour. Res., 28(4): 1089–1093.

Rodriguez-Iturbe, I. Rinaldo, A., Rigon, R., Bras, R.L. and Marani, 1991b. Energy dissipation, Runoff Production, and the 3-dimensional structure of river basins. Water Resour. Res., 28(4): 1095–1103.

Scheidegger, A.E., 1970. Theoretical Geomorphology. Springer Verlag, Berlin.

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

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

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

Smart, J.S., 1978. The analysis of drainage network composition. Earth Surf. Processes, 3: 129–170.

Smith, K.G., 1950. Standards for grading texture of erosional topography. Am. J. Sci., 248: 655–668.

Smith, T.R. and Bretherton, F.P., 1972. Stability and the conservation of mass in drainage basin evolution. Water Resour. Res., 8(6): 1506–1529.

Solow, A.R., 1987. Testing for climate change: an appli-

cation of the two phase regression model. J. Climate Appl. Meteorol., 26: 1401–1405.
Strahler, A.N., 1952. Hypsometric (area–altitude) analysis of erosional topography. Geol. Soc. Am. Bull., 63: 1117–1142.
Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Trans. Am. Geophys. Union, 38: 913–920.
Tarboton, D.G., 1989. The analysis of river basins and channel networks using digital terrain data. Sc.D. Thesis, Department of Civil Engineering, M.I.T., Cambridge, MA, (Also available as Tarboton D.G., R.L. Bras and I. Rodriguez-Iturbe, (Same title), Technical report no 326, Ralph M. Parsons Laboratory for Water resources and Hydrodynamics, Department of Civil Engineering, M.I.T., September 1989).
Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1989. Scaling and elevation in river networks. Water Resour. Res., 25(9): 2037–2051.
Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1990. Comment on "On the Fractal Dimension of Stream Networks" by Paolo La Barbera and Renzo Rosso. Water Resour. Res., 26(9): 2243–2244.
Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1991. On the extraction of channel networks from digital elevation data. Hydrol. Processes, 5(1): 81–100.
Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1988. The fractal nature of river networks. Water Resour. Res., 24(8): 1317–1322.
Thornes, J.B., 1983. Evolutionary geomorphology. Geography,: 225–235.
Tokunaga, E., 1978. Consideration on the composition of drainage networks and their evolution. Geographical Reports of Tokyo Metropolitan University, No. 13.
Willgoose, G.R., 1989. A physically based channel network and catchment evolution model. Ph.D. Thesis, Civil Engineering, M.I.T., Cambridge, Mass.
Willgoose, G.R., Bras, R.L. and Rodriguez-Iturbe, I., 1990. A model of river basin evolution. EOS, 71(47): 1806–1807.
Willgoose, G.R., Bras, R.L. and Rodriguez-Iturbe, I., 1991a. A coupled channel network growth and hillslope evolution model 1. Theory. Water Resour. Res., 27(7): 1671–1684.
Willgoose, G.R., Bras, R.L. and Rodriguez-Iturbe, I., 1991b. A coupled channel network growth and hillslope evolution model 2. Nondimensionalization and applications. Water Resour. Res., 27(7): 1685–1696.
Willgoose, G.R., Bras, R.L. and Rodriguez-Iturbe, I., 1991c. A physical explanation of an observed link area–slope relationship. Water Resour. Res., 27(7): 1697–1702.
Wolman, M.G., 1955. The natural channel of Brandywine Creek, Pennsylvania. U.S. Geol. Surv., Prof. Pap., 271, 56 pp.